From DEPARTMENT OF NEUROBIOLOGY, CARE SCIENCES AND SOCIETY, DIVISION OF PHYSIOTHERAPY
Karolinska Institutet, Stockholm, Sweden

INTERDISCIPLINARY REHABILITATION IN PATIENTS WITH CHRONIC PAIN
Prognostic Factors and Effectiveness

Elena Tseli

Karolinska Institutet

Stockholm 2019

Published by Karolinska Institutet.
Printed by Eprint AB 2019

ISBN 978-91-7831-510-9

INTERDISCIPLINARY REHABILITATION IN PATIENTS WITH CHRONIC PAIN: PROGNOSTIC FACTORS AND EFFECTIVENESS

THESIS FOR DOCTORAL DEGREE (Ph.D.)

By

Elena Tseli

Principal Supervisor:
Associate professor Björn Äng
Karolinska Institutet
Department of Neurobiology,
Care Sciences and Society
Division of Physiotherapy

Co-supervisor(s):
Associate professor Wim Grooten
Karolinska Institutet
Department of Neurobiology,
Care Sciences and Society
Division of Physiotherapy

Professor Björn Gerdle
Linköping University
Department of Medical and Health Sciences

Opponent:
Professor Pernilla Åsenlöf
Uppsala University
Department of Neuroscience
Division of Physiotherapy

Examination Board:
Professor Anne Söderlund
Mälardalen University
School of Health, Care and Welfare

Associate professor Carl Molander
Karolinska Institutet
Department of Clinical Sciences,
Danderyd Hospital

Associate professor Iben Axén
Karolinska Institutet
Institute of Environmental Medicine
Intervention and implementation research for worker health

With all my love to my beautiful little family,

Ioannis and Giorgos

ABSTRACT

Interdisciplinary multimodal pain rehabilitation (IMPR) is currently considered best practice for combatting chronic pain. However, it is believed that health-related outcomes could be improved with more adequately tailored treatment programs, but consensus of what grounds these adaptations should be based on is yet to be reached. Well-powered evaluations of naturalistic, real-world practices provide an evidence base for the evaluation of important characteristics that may facilitate the informed development of IMPR. The aim of the present work was to meta-synthesize existing evidence and add new data to the body of published evidence on prognostic factors for a positive outcome in patients receiving rehabilitation for chronic pain. An additional aim was to evaluate the effectiveness of different IMPR program durations on health-related quality of life in this major patient group.

Methods: Published international evidence of prognostic factors for physical functioning after IMPR was evaluated through a systematic review and meta-analyses (*Study* I), followed by the investigation of the inter-rater reliability of the Quality in Prognostic Studies tool (QUIPS), used in the Risk of Bias assessment (*Study* II). Prognostic factors (*Study* III) and effectiveness (*Study* IV) of Swedish pain specialist IMPR on *physical and mental functioning* and related measures of disease impact were investigated using large-scale nationwide data obtained from the Swedish Quality Registry for Pain Rehabilitation.

Results: Meta analyses showed, with moderate to low levels of evidence, that better physical functioning at follow-up was predicted by high levels of self-reported functioning, low levels of emotional distress and cognitive-behavioral risk factors, and high levels of cognitive-behavioral protective factors. Pain-related factors (intensity and chronicity) were not associated. Weak to moderate inter-rater agreement emerged for QUIPS, and suggestions for improving the inter-rater agreement and functionality were presented. Swedish registry data showed the most important prognostic factors were retaining a connection with work, having high optimistic treatment expectations, sense of control, and less interference from pain. Pain itself was of secondary significance. Also for improvement of physical functioning, better initial mental wellbeing was of importance, while for mental functioning the opposite emerged. Results on within-group effectiveness showed improvements on all outcomes, while no between-group comparison emerged on short (4-9 wks) *vs.* moderate (10 wks) *vs.* long (11-18 wks) IMPR program duration.

In summary, evidence for prognostic factors was identified, providing suggestions for the targeting of modifiable factors in clinics and in future clinical trials. Clearly, the quality assessment of published results needs systematic consensus work between assessors. Work connection, treatment expectations, levels of physical and emotional health, and coping strategies played an important prognostic role but were not consistent for physical and emotional functioning, suggesting a complex prognostic picture for the overall understanding of improvement. Finally, IMPR is effective across a biopsychosocial specter, but treatment duration seems not to play an important role.

SVENSK SAMMANFATTNING

Multimodal smärtrehabilitering (MMR) anses vara bästa praxis för behandling av kronisk smärta. De påvisade effekterna på olika (biopsykosociala) hälsorelaterade resultat är dock suboptimala. Det antas att rehabiliteringsresultat kan förbättras med mer adekvat anpassade behandlingsprogram, men fortfarande saknas konsensus om vilka faktorer dessa anpassningar bör grundas på. Effektiva utvärderingar av reell klinisk verksamhet kan ge en grund för utvärdering av viktiga egenskaper som kan bidra till utveckling av MMR. Syftet med detta avhandlingsarbete var därför att meta-syntetisera befintlig evidens, och att komplettera med nya primärdata till kunskapsunderlaget om prognostiska faktorer för ett positivt behandlingsutfall hos patienter med kronisk smärta som genomgått rehabilitering. Syftet var vidare att utvärdera effektiviteten av MMR som genomförs med olika behandlingslängd, med avseende på hälsorelaterad livskvalitet i denna omfattande patientgrupp med kronisk smärta.

Publicerade internationella resultat vad gäller prognostiska faktorer för *fysisk funktion* efter MMR utvärderades genom en systematisk litteraturgranskning och meta-analyser (*Studie I*). Denna studie följdes av en undersökning av inter-bedömarreliabilitet (tillförlitlighet) för ett nytt kvalitetsbedömnings-instrument avsett för prognostiska studier- ”QUIPS”, ett instrument som används vid bedömning av Risk för Bias (*Studie* II). Prognostiska faktorer (*Studie III*) och effektivitet (*Studie IV*) vid svensk MMR (specialistnivå) med avseende på *fysisk* och *psykisk funktion*, samt konsekvenser av smärta, utvärderades med hjälp av omfattande registerdata från Nationella Registret över Smärtrehabilitering, NRS.

Metaanalyser visade, med måttlig till låg evidens, att bättre fysisk funktion vid uppföljning förutspåddes av högre nivåer av självrapporterad funktion, låga nivåer av mental ohälsa, och kognitiva beteende-riskfaktorer samt höga nivåer av skyddande kognitiva beteende-faktorer. Smärtrelaterade faktorer (intensitet och duration) var inte associerade. Överenstämmelse vid Risk för Bias-bedömningar med QUIPS var svag till måttlig, och förslag till att förbättra överrensstämmelse och funktionaliteten presenterades. Svenska registerdata visade att de viktigaste prognostiska faktorerna generellt var att vara i arbete, ha optimistiska behandlingsförväntningar, känsla av kontroll och mindre funktionsstörningar från smärta. Smärta i sig var av sekundär betydelse. För *fysisk funktion* specifikt var ett gott mentalt välbefinnande av betydelse medan för *mental funktion* var ett lägre (dåligt) utgångsvärde av betydelse. MMR som helhet var effektiv vid behandlingsuppföljning för alla utfallsvariabler. Ingen skillnad framkom dock vid jämförelse mellan grupper med olika behandlingstid.

Sammanfattningsvis identifierades evidens för att det finns viktiga prognostiska markörer, vilket pekar på modifierbara faktorer användbara i såväl klinisk verksamhet som för framtida randomiserade kliniska studier. Det är tydligt att kvalitetsbedömning av publicerade resultat kräver systematiskt konsensusarbete mellan bedömare. Att vara i arbete, ha positiva behandlingsförväntningar, samt lägre ingångsvärden på de respektive utfallen, fysisk och emotionell hälsa, samt copingstrategier, spelade en viktig roll för det framtida

behandlingsresultatet. Resultaten var dock inte konsekventa för fysisk och emotionell funktion, vilket antyder en komplex prognostisk bild för förståelsen av klinisk förbättring som helhet. MMR är en effektiv behandlingsmetod från ett biopsykosocialt perspektiv, men dess behandlingslängd verkar inte ha någon betydande roll för det framtida behandlingsresultatet.

LIST OF SCIENTIFIC PAPERS

I. Tseli E, Boersma K, Stålnacke B-M, Enthoven P, Gerdle B, Äng BO, Grooten WJA. **Prognostic factors for physical functioning after multidisciplinary rehabilitation in patients with chronic musculoskeletal pain. A systematic review and meta-analysis.** *Clinical Journal of Pain 2019;35(2):148-73.*

II. Grooten WJA, Tseli E, Äng BO, Boersma K, Stålnacke B-M, Gerdle B, Enthoven P. **Elaborating on the assessment of the risk of bias in prognostic studies in pain rehabilitation using QUIPS—aspects of interrater agreement.** *Diagnostic and Prognostic Research. 2019;3(1):5.*

III. Tseli E, Vixner L, Lo Martire R, Grooten W, Gerdle B, Äng BO. **Prognostic factors for improved physical and emotional functioning 1 one year after interdisciplinary rehabilitation in patients with 2 chronic pain: results from a national quality registry (SQRP).** *Submitted for publication.*

IV. Tseli E, Lo Martire R, Grooten W, Gerdle B, Vixner L, Äng BO. **What is the effectiveness of different duration interdisciplinary treatment programs in patients with chronic pain? A large-scale longitudinal register study.** *In manuscript.*

Study III - IV may not be the final version for publication.

CONTENTS

LIST OF ABBREVIATIONS

EQ-5D	EuroQol-5 dimensions
GEE	Generalized estimating equation
GRADE	Grading of Recommendations, Assessment, Development and Evaluation
HADS	Hospital Anxiety and Depression Scale
HRQoL	Health-Related Quality of Life
IASP	International Association of the Study of Pain
ICD	International Classification of Diseases
ICF	International Classification of Functioning, Disability, and Health
IMMPACT	Initiative on Methods, Measurement, and Pain Assessment in Clinical Trials
IMPR	Interdisciplinary multimodal pain rehabilitation
MCID	Minimal clinical important difference
MDR	Multidisciplinary rehabilitation
MMR	Multimodal Rehabilitation
MPI	Multidimensional Pain Inventory
NRS	Numeric Rating Scale
OR	Odds Ratio
PCA	Principal component analysis
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PROM	Patient Reported Outcome Measure
PROSPERO	International Prospective Register of Systematic Reviews
QUIPS	Quality In Prognosis Studies
RCT	Randomized controls trials
RoB	Risk of Bias
SBU	Swedish Agency for Health Technology Assessment
SF-36	The 36-Item Short Form Health Survey
SF-36 MCS	SF-36 Mental Component Summary
SF-36 PCS	SF-36 Physical Component Summary
SQRP	Swedish Quality Registry for Pain Rehabilitation

VAPAIN	Validation and Application of a patient relevant core outcome set to assess effectiveness of multimodal PAIN therapy

1 INTRODUCTION

"*My pain never takes time off, it has a will of its own*"

"*My pain and I now live together. Not always in harmony*."

Chronic pain conditions result in wide-ranging negative effects on a person's life and are associated with a great amount of suffering. Health-related quality of life in patients with chronic pain has actually been reported to be among the lowest of any medical condition. In addition to discomfort and distress for the affected individuals, chronic pain imposes a considerable economic burden on the health care system and society as a whole and, despite all efforts, it is increasing – chronic pain is now regarded as an epidemic condition.

Interdisciplinary multimodal pain rehabilitation (IMPR) is currently the preferred treatment for chronic pain. It was first described in the 1970s, has been continuously developed since then, and has now solidified its position within the health care systems of the Western world. In Sweden, approximately 40 clinics specializing in the treatment of chronic pain offer IMPR (in Sweden commonly referred to as Multimodal Rehabilitation, MMR) to patients with significant levels of persistent pain. Current research reports the effectiveness of IMPR to be only moderate, but despite this it is considered to be the best practice available today. There is, however, a great need to further increase knowledge regarding the selection and adaptation of treatment so it better matches patients' needs, especially since IMPR is costly and patients are selected for different treatment approaches based on limited evidence and local clinical preferences. There is a need for evidence built on larger sample populations and a higher number of clinics in order for a representative evaluation of pragmatic practice to be possible. This, in turn, could further indicate areas for future detailing with experimental designs.

This thesis was designed with the long-term general goal of helping improve health care in general and specific goal of helping individuals with chronic pain in the health care system. The project adopts a longitudinal design where IMPR is prospectively evaluated with respect to pain, functioning, and health-related quality of life. It addresses the challenge of combating chronic pain and constitutes a key passage to a translational process, and is designed to evolve current knowledge through realistic results derived from practice-based evidence. It includes four studies; the first is based on data from published original studies included in a systematic review and meta-analysis, the second is based on data from method evaluation as part of the system review, and the remaining two are founded on nationwide data from the Swedish Quality Registry for Pain Rehabilitation (SQRP). By using a large sample approach founded on both national and international multicentre data, it is hoped that this study will contribute to existing evidence on prognostic factors and the effectiveness of IMPR – currently the best available, yet still suboptimal, treatment for chronic pain.

1.1 PERSPECTIVE AND THEORETICAL FRAMEWORK

This thesis is based on a positivistic perspective, with a deductive approach applied to the research questions, as often seen in quantitative research. The project investigates an interdisciplinary research field, and as such expands to broadly encompass rehabilitation medicine and public health. Figure 1 gives an overview of the thesis topic, research questions, and used data sources.

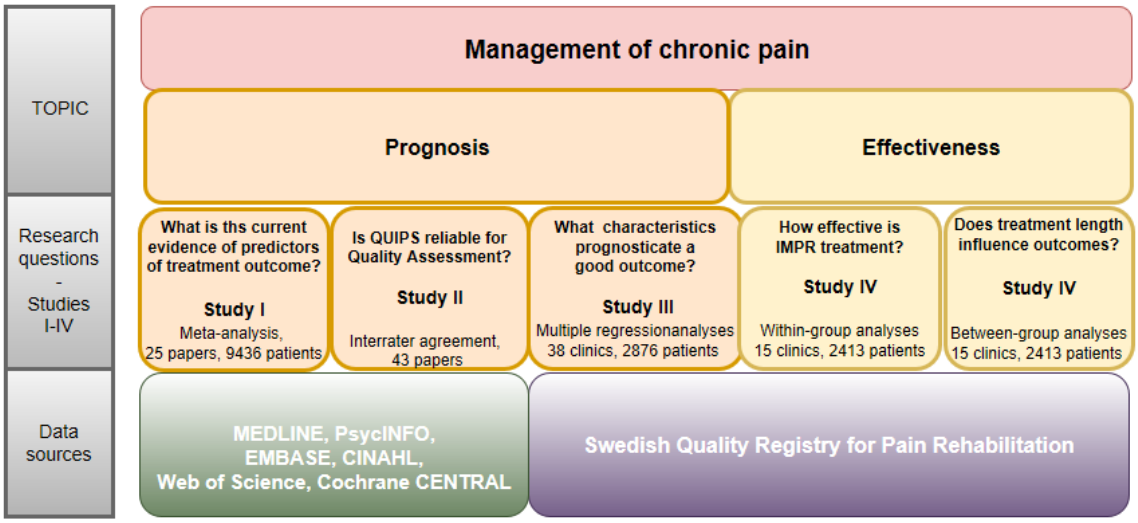


Figure 1. Overview of topic, research questions and data sources of the thesis.

The present thesis addresses interdisciplinary rehabilitation as a way to improve health in people with chronic pain, making the concept of health central. A commonly used definition of health, endorsed here, is from the World Health Organization (WHO) in 1948: "Health is a state of complete physical, mental and social wellbeing and not merely an absence of disease or infirmity", later complemented with "the ability to lead a socially and economically productive life" (1, 2). Health here is, therefore, described as a bio-psychosocial concept, which also involves contextual factors. The health model of the International Classification of Functioning, Disability, and Health (ICF) can be used as a reference for the description of a person's health condition and provides a language to depict consequences of this health condition. The overall term in the framework is *Functioning*, which covers all body functions, activities, and participation, while *Disability*, on the other hand, is an overall term for impairments, activity limitations, and participation restrictions. The model also addresses interacting contextual factors that are both personal and environmental (3). The ICF model thus reflects the multidimensionality of the health concept, however, it does not depict the equally central concept of health-related quality of life (HRQoL), which is addressed in this thesis. It has been suggested that the dimension quality of life as concept can be imagined as the 'circle around the ICF-model'(4). In HRQoL, perceived physical and mental health are believed to constitute core dimensions (5), and therefore in this thesis, the term is used as an umbrella term for our main outcomes.

2 BACKGROUND

2.1 PAIN - DEFINITIONS AND TERMINOLOGY

2.1.1 Pain

The International Association of the Study of Pain (IASP) endorses the definition of pain as "an unpleasant sensory and emotional experience associated with actual or potential tissue damage, or described in terms of such damage" (6). This definition emphasizes that pain refers to an experience and is, therefore, always subjective. It is "unquestionably a sensation in a part or parts of the body, but it is also always unpleasant and therefore also an emotional experience" (7). As such, the experience of pain is inherently multidimensional by nature. Still, a criticism to the above definition is that other major dimensions of the pain experience are excluded, which are equally important from a clinical perspective. Probably there is no perfect definition for the abstract and subjective phenomenon of pain, but the following suggested additional dimensions are of great relevance, in particular where chronic pain is concerned: "Pain is a distressing experience associated with actual or potential tissue damage with sensory, emotional, cognitive, and social components."(8).

2.1.2 Chronic pain

Several starting points for the classification of pain exist, e.g. temporal, etiological, or mechanism-based. When pain persists and exceeds, a pre-defined temporal cut-off is commonly used and pain lasting for a period of ≥3 months, or beyond the point of normal tissue healing, is referred to as persistent or chronic pain (6). However, classifying chronic pain by duration solely has some shortcomings from a clinical perspective. Moreover, the vocabulary used for common chronic pain conditions is not very consistent; sometimes it is based on the anatomical location of pain (for instance chronic low back pain, neck pain, or widespread pain), or on the pain mechanisms involved (nociceptive, inflammatory, neuropathic, and recently nociplastic pain) (9). More focus on identifying the underlying pain mechanisms is believed to be very important for the pain management selection process (10).

It has long been argued that although etiology, localization, and diagnoses might differ, chronic pain itself could be considered "a disease in its own right" (11). Hence, in view of the 11th revision of the International Classification of Diseases (ICD) (12, 13), the application for a separate diagnostic code for chronic pain in the updated classification system is intended to better reflect the actual epidemiology of chronic pain conditions, which will in turn better facilitate their clinical management.

Chronic primary pain is defined as pain in one or more anatomical regions that (1) persists or recurs for longer than 3 months, (2) is associated with significant emotional distress (e.g.

anxiety, anger, frustration, or depressed mood) and/or significant functional disability (interference to activities of daily life and participation in social roles), and (3) the symptoms are not better accounted for by another diagnosis.

2.1.2.1 *Chronic musculoskeletal pain*

Chronic musculoskeletal pain represents the most prevalent set of chronic pain conditions and includes the vast majority of chronic pain conditions that are the focus of this thesis. In accordance with the new ICD-11 definition of chronic pain, the majority of these non-malignant and non-systemic disease pain conditions would be classified as *chronic primary pain (12)* or *chronic musculoskeletal pain* (primary, secondary, or nonspecific) (14). In this thesis, the term "chronic musculoskeletal pain" will be referred to as synonymous with the term "chronic pain".

2.2 CHRONIC PAIN

2.2.1 Epidemiology

The World Health Organization (WHO) recognizes chronic pain as the leading global public health problem with 27% of the adult population living with chronic pain of significant intensity (15). Of all the chronic pain conditions, musculoskeletal diseases are predominant, prevalent in nearly one third of the population (16). This type of chronic pain is the most common cause of long-term disability in middle-aged people (17). Years lived with disability (YLDs) is a measure of non-fatal health outcomes, and pain conditions have been found to cause 21% of all YLDs globally. In a Lancet study published in 2017 (18), low back pain emerged as the leading single cause for YLDs followed by major depression disorders, anemia, and neck pain. These leading causes were significant when scaled to Europe and Sweden as well. The problem also seems to be increasing in both Sweden and worldwide. Chronic pain is also reportedly associated with an increased risk of mortality, even when controlled for socio-demographic factors (19, 20). Taken together, chronic pain presents a major public health problem with major health cost implications for society (15).

2.2.2 The societal burden of chronic pain

Chronic pain causes a considerable economic burden on society that significantly affects every country's GDP. Chronic pain is also associated with a high degree of health care consumption (15) and frequently results in long-term absences from the workplace (>180 days). This constitutes major socioeconomic problems not only for individuals but also for health care systems and societies as a whole. In Sweden, the costs for chronic pain in the form of social insurance costs for sick leave and loss of production have been estimated by the Swedish Agency for Health Technology Assessment (SBU) to be 90 billion Swedish kronor (90,000,000,000 SEK) annually (21). And the cost of musculoskeletal disorders represented

31% of total health costs in Sweden, the second largest expenditure on healthcare in Sweden, exceeded only by costs associated with all mental illnesses together (22).

2.2.3 Etiology

The mechanisms that underlie a transition of acute musculoskeletal pain into chronic pain are only partially uncovered, and although huge advances in the field have been made in recent years, the etiology behind why some people develop chronic pain and others do not still remains unclear (23, 24). Acute musculoskeletal pain is often triggered by injury or the overload of musculoskeletal structures (muscles, tendons, ligaments, joints, and skeletal structures) causing inflammatory, ischemic, or degenerative processes (25) that result in nociceptive pain (25). Low back pain, neck pain, and myalgia, for instance, are common benign musculoskeletal disorders that often occur and, most commonly, the initial acute pain associated with them will decline in due time as healing occurs (23).

Taken together, advances in the field of neurobiology have led to a wider understanding of chronic pain, but the exact underlying pathophysiology is still undetermined. In the past 50 years, knowledge of mechanisms in pain transfer and modulation has increased rapidly. Milestones such as the Gate Control Theory in 1965 (26), the Neuromatrix Theory of Pain in 1999 (27), and more recent advances in pain physiology and imaging techniques, have all provided evidence of complex mechanisms acting within the nervous system; how alterations occur that may contribute to the experience of pain by activating pain memories in the absence of peripheral input, or by exaggerating noxious or non-noxious inputs. This explains how neurobiological changes can occur even in the absence of injury or inflammation, which can partly explain the mechanisms of persistent pain. Central sensitization is considered to play an important role in chronic pain conditions and refers to neural alterations that may perpetuate and heighten the pain signal. Some of the pain facilitating mechanisms are a) increased reactivity (sensitization) in the peripheral and central nervous system, e.g. wind up, b) increased sensitivity to pain, e.g. hyperalgesia and allodynia, and c) decreased (dysfunctional) endogenous pain inhibition (28, 29). A complex functional network, the Neuromatrix, and its related modulations by somatosensory, endocrine, immunological, and autonomous systems, is activated in pain, and emotional and cognitive processes influence both inhibitory and facilitatory processes (30).

2.2.3.1 *Risk factors for developing chronic pain*

The complex and multifactorial nature of pain challenges knowledge of set risk factors for pain persistence and development of chronic pain conditions. High initial pain intensity that markedly affects everyday life and long pain duration are suggested risk factors. Several other factors of importance have come forth as well, displaying a biopsychosocial panorama; personal characteristics e.g. advanced age, female sex, non-native ethnicity, comorbidity and

self-reported low health, psychological factors such as high psychological loading, inadequate coping strategies and avoidance behavior, socio-demographic and work related factors including high physical demands and a negative psychosocial environment, and long periods of sick leave (28, 31-34). Hence, a wide range of possible risk factors has been proposed and amongst these, psychosocial factors are believed to play a greater role along with increasing chronicity. Efforts to identify these early, for instance through specific questionnaires, seem important for efforts made to avoid transitioning from an acute to a chronic problem (35). Several biological, psychological, and socio-demographic components may have an impact on this process, which necessitates the biopsychosocial approach for this multifactorial phenomenon (34).

2.2.4 A biopsychosocial understanding of a complex condition

Chronic pain is a complex condition, differing from acute pain in more ways than simply duration. A multifaceted and dynamic interaction between physiological, psychological, and social factors are involved in numerous processes that perpetuate, and even strengthen, one another resulting in complex conditions with chronic pain and disability (30). To understand and treat chronic pain, a biopsychosocial approach is regarded as fundamental, and this should be reflected in all meetings with the patient throughout treatment – from the first encounter, to the assessment, planning and implementation of intervention as well as in evaluating treatment effects (30). The biopsychosocial model, initially proposed by GL Engel in 1977 (36) was introduced to expand the traditional biomedical perspective which was too narrow to hold the complexity of the condition and hence hindered an optimal treatment approach (30). The model has been increasingly influential in the field of pain, especially as a model to understand chronic pain.

Pain, its interference in most daily activities and decreased functioning are common consequences of chronic pain resulting in physical deconditioning over time (37). When pain lasts for a longer period of time, emotional and behavioral components have more opportunity to interact with the social environment. Pain generates frustration and fear, affecting both the person afflicted and his/her environment. This psychological distress may result in anxiety and depression disorders and, in addition to the emotional agony, it is not uncommon for sufferers to experience additional stress, such as financial stress from reduced work ability, marital and family dysfunction, substance abuse and dependence, and a general decline in social and recreational functioning (38). Chronic pain results in social and psychological consequences and affects work, leisure, future perspectives, and self-image. Pain, in addition to its consequences, may lead to inhibited participation in society. Marginalization from work, alienation within families and relationships, and otherwise restricted engagements in active everyday life are frequent. Still, the impact and extent of the consequences vary, some may have more limited negative consequences while others experience severe pain and overwhelming distress (39, 40).

Sufferers of chronic pain score one of the lowest HRQoL scores, equal to that of patients with advanced cancer in palliative care (41). The psychosocial aspects of pain were the predominant reason for this rather than the pain intensity itself. The overall picture of consequences of chronic pain hence extends over all dimensions of an individuals' functioning, activity, and participation as depicted in the ICF health model, and are influenced by personal as well as environmental factors (42). Hence, any efforts to treat such a complex condition must naturally encompass these dimensions in therapy in order to be successful, and therefore complex interventions have gained acknowledgement in the field of pain rehabilitation.

2.3 INTERDISCIPLINARY MULTIMODAL PAIN REHABILITATION

Rehabilitation medicine forms the extension of curative medicine and specializes in treating complex medical conditions accompanied by longstanding disability. Three rehabilitation procedures have been described for chronic pain, depending on the amount of measures applied; 1) Unimodal rehabilitation, where one single modality is applied, e.g. physiotherapy and the intervention is performed by healthcare workers from one profession, although it may include different approaches (43). 2) Intermediate rehabilitation, where more measures are applied by healthcare workers from different professions who may convene in teams but provide their treatment independently, and 3) Interdisciplinary multimodal pain rehabilitation (IMPR), which is the most complex intervention involving a set team of several healthcare workers representing a number of professions, and with the patient being an active member of the team (28, 44).

2.3.1 Definition of Interdisciplinary multimodal pain rehabilitation

There is a vast use of different terminology used in both literature and clinical practice for these complex treatment approaches, which is why the IASP recently proposed using unified terminology, with the term interdisciplinary treatment to be used for: *"Multimodal treatment provided by a multidisciplinary team collaborating in assessment and treatment using a shared biopsychosocial model and goals."* A similar expression, which also is being increasingly used, is Interdisciplinary Multimodal Pain Therapy (IMPT) (45). In Sweden, IMPR is often referred to as multimodal rehabilitation (MMR) or multidisciplinary rehabilitation (MDR). The term IMPR will be used in this thesis and is synonymous with these Swedish terms.

2.3.2 Therapeutic targets, composition, and administration of IMPR

IMPR is directed towards all the diverse components of the pain experience as well as their consequences. It targets functioning and the psychosocial dimensions of chronic pain, and, by extension, health-related quality of life. Central features are based on behavioral medicine and theories on learning, using a cognitive behavioral therapy approach to address patient

cognitions, emotions, and behaviors with the aim to improve functioning and coping ability. Its approach considers each patient's needs from a bio-psycho-social perspective, including work or employment status. IMPR generally includes patient education and supervised physical activity/exercise therapy as one of the core treatment options along with psychological measures, occupational therapy sessions, ergonomics, and pharmaceutical treatments if needed (46, 47). IMPR is administered by specialized interdisciplinary teams, usually consisting of physicians, psychologists, physiotherapists, occupational therapists, social workers, and other healthcare professionals (21).

A rehabilitation plan is formed in collaboration with the patient. Most interventions are group-based but each patient has an individual plan, schedule, and individual contacts according to his/her own goal setting (28, 48). IMPR-measures are expected to act both on their own (21, 47) and in conjunction with each other, equivalent to what has been defined as a complex intervention by Medical Research Council guidance (49). The sum of these parts is supposedly greater than if they were provided unimodally.

IMPR dosage is an equally complex matter to delineate. Evidence is still unclear on whether more or less optimal administrations of IMPR exist with regard to time aspects, such as how long a treatment program should be in order to be effective and if intensity should be viewed in this context. Comparisons of the composition, intensity, and duration of IMPR programs remain inconclusive (50). Principles of behavior change imply that behavioral changes require time to learn and practice and therefore interventions should provide sufficient time to adopt these skills into daily life (51, 52). Healthcare professionals' experience also add to that treatment length is of importance for learning and implementation to take place (53). However, until now, there has been no available empirical evidence of the duration of treatment programs having an influence on the success of the intervention (47, 50).

2.3.2.1 *IMPR in Sweden*

IMPR is currently offered by approximately 40 specialist clinics throughout Sweden. Patients of working age with benign chronic pain from the musculoskeletal system who experience complex adverse effects from this in their lives can be referred to participating clinics for investigation and/or interdisciplinary rehabilitation. The referrals to participating clinics come primarily from primary care and private general practitioners, but also from other clinics and hospitals.

Patients are initially assessed by a multidisciplinary team through interviews and examinations, and patients with potential "red flags", such as history of trauma, unexplained weight loss or a cancer history, are excluded before a rehabilitation plan is formed. With a focus on increasing accessibility and effectiveness, two levels of IMPR treatment plans have been set; MMR 1 for less complex pain conditions, mainly provided in primary care, and MMR2 for extensive and more complex pain conditions, primarily provided by specialist clinics. All of Sweden's

specialist clinics are linked to a quality registry, the Swedish Quality Registry for Pain Rehabilitation (SQRP) for the purpose of treatment evaluation and development.

2.3.3 Outcomes in rehabilitation of chronic pain

Pain, physical functioning, and emotional functioning are core outcome domains suggested by the Initiative on Methods, Measurement, and Pain Assessment in Clinical Trials (IMMPACT) group (54), along with participant satisfaction with treatment and symptoms or adverse events. The importance of evaluating multiple outcome domains was confirmed consistently across pain conditions in a large survey of people with chronic pain (55), and it has therefore been suggested that it is included in the study of the efficacy/effectiveness of chronic pain treatments. A variety of appropriate measures for these patient-reported outcome measures exist, commonly referred to as PROMs (Patient-Reported Outcome Measures) and PREMs (Patient-Reported Experience Measures). They adhere to the multidimensionality and subjective reporting of the experience of pain, as emphasized by the IASP definition of pain. The majority of targeted outcomes are thus assessed by PROMs, both regarding pain, and physical and emotional functioning. Physical functioning is an important outcome domain, as patients experience major restrictions due to functional problems which extend across all domains of the ICF model. Physical functioning is also a broad construct, and can be measured in several ways; performance based or through self-reporting, but predominantly in the field of IMPR it is measured though PROMs. There is an abundancy of measures, some measure pain-related physical functioning specifically, while others measure general physical functioning. When evaluating treatment benefits related to physical functioning and functioning in general, one can assess either proximal aspects of functioning, closely related to the core signs and symptoms of the disease, or more distal and general disease impact concepts such as general physical, psychological, and social functioning and, by association, general life concepts such as health-related quality of life (56).

2.3.4 Prognosis in IMPR

Prognosis refers to a likely future outcome; in this context it refers to the likely health outcomes of people who participate in IMPR. Prognostic factor research aims to identify factors associated with better or worse outcomes, in order to inform therapeutic decisions and help identify interventional targets that may help modify the likely course of the condition (57). Knowledge of factors of importance for the prognosis of treatment outcome is vital but greatly limited, and in the field of IMPR the body of evidence is still insufficient to provide any guidance. It is currently unclear which factors, patient related and/or treatment related, are essential for treatment success, and which indicate a risk for a poor prognosis (21). Previous reviews investigating prognoses for a range of core outcomes on diverse chronic pain, fibromyalgia (58), or chronic low-back pain populations (59, 60) have identified some patient

characteristics (i.e. prognostic factors) including patients' initial levels of depression, initial pain intensity, and work-related functioning and active coping skills.

The prevailing evidence, however, is mainly based on narrative syntheses, as the included studies were considered too heterogeneous for data to be pooled for a meta-analysis. The authors of systematic reviews on IMPR outcome predictors generally conclude that studies are of low quality, often using a simple exploratory design, making the identification of a universally valid set of predictors impossible. In general, combinations of predictors seem more important than single predictors, in accordance with chronic pain being a complex multifactorial problem. The SBU and national networks acknowledge the importance of acquiring a better evidence base for prognosis, in order to gain guidance on selecting and customizing the best possible invention for each individual case.

2.3.5 Effectiveness in IMPR

IMPR for chronic pain has been evaluated and, as indicated, is generally considered more effective than unimodal or less complex interventions (21, 47, 48, 61-63). Conducting high-quality research into IMPR has been recognized as a major challenge due to the complexity of the conditions and diverse nature of the IMPR programs, with respect to content and intensity (or dosage as defined herein). To date, there is little scientific literature on how IMPR interventions should be designed to optimize results (45). It is still unknown which treatment components are really important, and whether all components benefit patients alike (47). This lack of evidence has delayed the improvement of existing IMPR programs. IMPR is expensive and time-consuming, and clinical departments actively select which patients are offered this intervention.

In addition to patient-related factors, other factors such as the specific orientation of the IMPR program and its dose may relate to the treatment itself and have interactions with different patient characteristics. Due to the limited number of studies investigating such variations, the effects of these still remain unclear, delaying evidence on whether some treatment modalities work better in some types of patients.

In an attempt to evaluate the effects of different IMPR intensities, which is one ingredient of dosage, Haldorsen et al. (64) found that patients with a good prognosis appeared to benefit from any intensity whereas patients with a poor prognosis benefitted only from extensive IMPR. However, their data suffered from selection bias as only 33% of those invited agreed to participate, and participants were recruited from only one outpatient clinic.

In addition to uncertainties regarding IMPR intensity, there is also a lack of knowledge regarding the optimal duration of IMPR programs, which range from four to 18 weeks in Sweden. Duration, which is another ingredient of dosage, seems to be often based on local tradition and medical staff preferences. IMPR duration is certainly complex and cannot be directly transferred from pharmacotherapeutic principles, but the topic has been extensively

discussed in pain rehabilitation networks and patient associations and has been acknowledged in the latest SBU report (21). The importance of program duration can be stressed since treatment strategies involving behavioral changes including adherence to physical exercise elements of a program take time in chronic pain conditions (52, 65). Preliminary data from a report on a limited sample indicate that a 12-week IMPR program is superior to a comparatively shorter IMPR program where health-related quality of life is concerned. However, further data from multiple centers is required, and the focus of future evaluations should also include cost-effectiveness, which has been requested for a long time (15).

In summary, there is an urgent need, called for by the healthcare community (4, 10), to find a basis for more adequately tailored rehabilitation programs in order to improve the treatment effects for both high-risk subgroups and for those with good prognoses. The need for more knowledge in the area has received attention due to IMPR being expensive and time-consuming (44).

2.4 THE CONCEPT OF EVIDENCE

The Swedish National Board of Health and Welfare (Swe: Socialstyrelsen) recommends that healthcare professionals use their *evidence-based practice* model. In this model, the choice of treatment is based on four cornerstones. First, one must reflect on the patient's clinical and physical circumstances as a start to find out what treatment options are available and what are not. Secondly, the selected treatment option should be strengthened by research evidence concerning its efficacy, effectiveness, and efficiency. Thirdly, if each treatment option is equally effective, the clinician must consider the patient's preferences and likely actions (in terms of what interventions she or he is ready and able to accept). Finally, it is necessary that clinical expertise is available, i.e. including both the clinician's knowledge and skills and the available resources in the health care system (66). While the first, third, and fourth pillar are very patient-specific, the second pillar is based on research evidence from scientific literature.

2.4.1 Research evidence through systematic reviews

Randomized controlled trials (RCTs) and pragmatic RCTs have been used as a golden standard to answer research questions regarding the efficacy and effectiveness of interventions, and since RCTs use the preferred methodology for causal inferences and due to randomization, they have the possibility to avoid confounding effects when comparing different treatment approaches. However, decisions about the usefulness of an intervention (or other types of research question) evidently cannot be based on the results of a single study – the evidence is too limited. Therefore, evidence synthesis is commonly based on systematic reviews of original studies, including RCTs as well as well-designed cohort studies, case-control studies, and sometimes even cross-sectional studies and pilot-studies(67).

In systematic reviews , a first step is to review the available literature narratively (or qualitatively), but this process is somewhat subjective since different experts can come to different conclusions, making drawing robust conclusions on the effectiveness of an intervention problematic when included studies are under-powered or contradictory. Re-analysis of data though meta-analyses , which use the quantitative aggregation of all data available, can instead increase power and make effects or associations apparent and objective.

Therefore, systematic reviews and meta-analyses are set on the top of the pyramid of evidence in evidence-based practice. In these processes, however, there is still a need to critically assess and judge the methodological quality of the studies they include. This process is called a Risk of Bias (RoB) assessment and there are several different tools developed for this purpose.

2.4.1.1 *Assessment of Risk of Bias*

Many systematic reviews and meta-analyses focus on synthesizing the results from RCTs, comparing outcomes from different interventions. However, comparisons are difficult to make when methodological issues such as attrition, response rate, intervention dosage, outcome measure, soundness of statistical methods, the internal and external validity, etc., differ between the included studies. There are several specific RoB tools for identifying methodological weaknesses in order to distinguish between studies with high risk of bias and studies with low risk of bias. In other types of studies, e.g. observational, cohort, or register studies, which are non-randomized prognostic studies, evidence could be provided by within-group comparisons on real-world data, which may give higher external validity. These studies are referred to as effectiveness studies, and require other types of RoB tools (68). For prognostic studies, a special RoB tool has recently been developed, and this tool has been suggested as a core tool in the performance of Cochrane review studies (69).

Since conclusions of the outcome of a systematic review or meta-analysis study are based on the methodological quality of the included studies, it is important that the RoB assessment itself is well performed and of high scientific quality. Therefore, in the most robustly designed systematic review and meta-analysis studies, two or more researchers will perform an RoB assessment independently from each other, after which they will discuss each study to reach a consensus on its methodological quality; i.e. if it has a high, medium, or low RoB. Hence, this important step in evidence synthesis is based on both the researchers' judgment and the instruments' ability to provide a clear basis for judgment, both of which may contribute to a potential 'measurement error'. Applying any measurement instruments also requires some evaluation of its measurement ability, here both referring to the researcher rating the instrument and the instrument itself. In this process, a calculation of the inter-rater agreement indicates aspects of reliability.

2.4.1.2 *Grading of Recommendations, Assessment, Development and Evaluation, GRADE*

As a final step to reach levels of evidence a GRADE is performed. The GRADE process aims to define the "*extent to which one can be confident that the estimate of effect or association is*

correct". It is a method that is used globally and that provides accurate estimates on specific outcomes that may be applied to recommendations for health care (70).

GRADE has four levels of evidence, ranging from "very low" (+) to "high" (++++). Depending on the research designs upon which the synthesis is based, the starting point is either regarded as being high quality, such as evidence from RCTs, while other less inferential designs are rated as being lower quality. From this initial point, evidence levels can be downgraded due to factors that may decrease confidence in the estimates, such as a high RoB or publication bias. Evidence levels can also be upgraded due to factors such as large magnitudes of effect.

2.4.2 Practice-based evidence

Although available evidence supports the effectiveness of IMPR, the body of evidence risks also being based on certain reports reflecting artificial situations with regard to selected patients, risk for researcher allegiance, and problems with practical applications. RCTs provide the foundation of clinical evidence to guide clinicians in their selection of treatment options, however, these trials have known limitations, including not only the immense costs associated with the implementation of adequately powered studies, but also limited generalizability to naturalistic practice settings with a consecutive non-selective flow of patients with no extra "RCT-resources" or specialists, commonly available in RCT trials (49). A necessary step is therefore to investigate whether the evidence reported in RCTs as well as systematic reviews also holds for a consecutive non-selective flow of patients in real-world practice settings.

Research methodology known as *practice-based evidence* evaluates the effectiveness of regimen programs through prospective observational cohort study designs, applied in rehabilitation research (71). The external validity in such designs is usually high because they include virtually all patients and potential confounders that could alter treatment responses (71). The methodology has certain characteristics (e.g. large sample sizes and multiple data sampling sites). Hence, both practice-based evidence and data gained through RCTs (or evidence-based practice) can be perceived as important complementary parts of an on-going translational "bench-to-bedside" research process to improve the outcomes of treatments for patients with chronic pain.

Pragmatic clinical registry-studies can be a powerful and highly cost-effective method of establishing clinical evidence of national relevance in areas that might otherwise be difficult to evaluate, such as rehabilitation research (71). Analysis on prospective quality registries commonly reflects regular clinical settings, and carries the additional advantage of not requiring a wait of several years to gauge long-term effects. Also, further openings for large observational databases have received increasing scientific interest, applying techniques to emulate a target trial for comparative effectiveness, when a randomized trial is not feasible (72). By identifying data that represents independent groups (e.g. short vs. extensive IMPR duration) in a large, clinical registry-cohort with consecutive enrolment, it is believed that

important features of a prospective randomized trial can be combined with the inclusiveness and efficiencies of a large-scale nationwide clinical registry.

2.4.3 Swedish National Quality Registries

The Scandinavian countries, including Sweden, have a tradition of keeping a high number of well-established national registries that provide unique opportunities to perform register-based research that can be linked to individual national registration numbers – personal identification numbers that every citizen has (73, 74). In addition to population-based registries run by central authorities, there is also a great amount of registries appointed for quality assessment of the healthcare system; Swedish National Quality Registries (75). Their data is collected in real time, i.e. prospectively, and integrated into clinical workflows. A group of dedicated stakeholders representing health care, patient, and research perspectives are jointly responsible for developing these registries.

Presently, about 100 National Quality Registries provide the Swedish healthcare system with a unique opportunity to monitor the quality of interventions and their results. Quality registries provide data that can be used in the evaluation of specific areas of healthcare, for instance heart disease, cancer, rheumatic diseases, etc. and include individualized data on patient problems, treatment, and outcomes usually followed over time. One of these is the Swedish Quality Registry for Pain Rehabilitation (SQRP).

2.4.3.1 *Swedish Quality Registry for Pain Rehabilitation (SQRP)*

The SQRP is a quality register that was founded in 1998 with its main aim being to promote research and development in the field of IMPR (76). The registry includes data from patients with chronic non-malignant pain, who undergo an evaluation for specialist pain management. A growing number of clinical units all over Sweden are linked to the registry and in 2017, about 40 specialist care units were included, giving it a coverage rate of 95% (geographic location is shown in Figure 2. Hence, the vast majority of patients in Sweden undergoing pain specialist care on tertiary level are included in SQRP. Every year, approximately 7000 new patients are added, representing at total of over 55,000 clinical visits (76). Recently, the SQRP was extended to include primary care units specializing in IMPR, where patients with less severe pain conditions and a lower psychological distress burden attend (77).

SQRP is based primarily on patient-reported health data of core outcome domains suggested by the IMMPACT for chronic pain studies (55) and the Validation and Application of a patient relevant core outcome set to assess effectiveness of multimodal PAIN therapy (VAPAIN) consensus statement (78), i.e. pain, physical and emotional functioning, quality of life. Data is collected using established questionnaires, both generic and disease specific, prior to and directly after treatment as well as after a one-year follow-up period. There is a minimum set of established PROMs that are used by all clinics, but these can be complemented with additional measures of the clinics' own preferences. Furthermore, data on diagnoses, socio-demographic details, information on employment, and sick leave status, as well as pain duration and future prospects, are collected.

The clinics, which routinely use the measures for the purpose of assessing individual patients, manually register data into the database of SQRP, provided they have the patient's permission. The general aim of SQRP is to compare patient groups at different rehabilitation clinics and the effect of rehabilitation programs according to patient-reported outcome measures concerning function, activity, quality of life, and participation in working life and leisure (75). The registry enables detailed analyses for evaluation of effectiveness and quality of care, and, through rigorous regulatory (and ethical) processes, data can also be linked to the Swedish Social Insurance Agency and other registries. Annual summary reports are provided from the SQRP and the database is also subject to large empirical research projects which are preceded by application processes with valid ethical permits.

2.5 RATIONALE FOR THESIS

Chronic musculoskeletal pain conditions constitute a major challenge for health care systems worldwide. Therapeutically, the necessity of a biopsychosocial approach using interdisciplinary rehabilitation is widely accepted – yet its effectiveness on several health outcomes continues to be only moderate. An important question is why do some patients undertaking IMPR improve, while others do not? What are their characteristics – is there any early information available from our national registries or from existing literature that can indicate how specific groups of patients will function after their rehabilitation? Regarding existing literature, a thorough evaluation of study quality of the studies upon which the evidence will be based is essential in systematic reviews and evidence synthesis. For the assessment of quality in prognosis studies, a new instrument that necessitates the evaluation of its reliability and functionality within this research topic has been introduced. What we know, however, is that existing evidence is inconclusive – and this hinders informed clinical decisions and the further development of IMPR. Clearly, many patients need support in managing pain and behavioral change that will lead to a healthier and more active life, but for how long do they need this support in organized IMRP? In Sweden, specialist IMPR programs range between four and 18 (or more) weeks in duration, and are often based on local tradition and medical staff preferences. As indicated, IMPR duration (one aspect of dosage) is certainly complex and cannot be directly transferred from pharmacotherapeutic principles. This topic has been extensively discussed in pain rehabilitation networks and patient associations and has been acknowledged in the latest SBU report on pain rehabilitation. Through the latest data in SQRP, now also including data on treatment duration, this thesis had the unique possibility to challenge the idea that "behavioral change takes time" by contrasting effects of IMPR duration on health-related outcomes. To the best of our knowledge, this is the first study of its kind that has a broad national representativeness.

Hence, the rationale for the present thesis was to provide updated evidence, that has been called for both nationally and internationally, on prognostic factors and the effectiveness of IMPR duration, in order to tackle the contemporary challenges with this major patient group.

3 AIMS

Given the existing gaps in knowledge, the overall aim was to meta-synthesize existing evidence on prognosis for a positive IMPR outcome, and to investigate the reliability and functionality of a new tool for RoB assessment that can be used in systematic reviews; the Quality in Prognosis Studies (QUIPS) tool. A further aim was to delineate the prognostic factors and effectiveness of IMPR, on dimensions of health-related quality of life by approaching large-scale nationwide data in patients with chronic pain receiving rehabilitation.

3.1 SPECIFIC AIMS

The specific aims of the studies included in the thesis were:

- to identify, evaluate, and meta-synthesize published data on prognostic factors, focusing on pain and physical and emotional functioning, for long-term follow-up physical functioning in patients with chronic pain *(Study I)*

- to determine the interrater agreement of the risk of bias assessment in prognostic studies using the instrument Quality in Prognosis Studies (QUIPS) and to elaborate on the use of this instrument, *(Study II)*

- to investigate prognostic factors for long-term physical and emotional functioning by targeting patients' baseline characteristics and health measures, *(Study III)*

- to evaluate the effectiveness of IMPR on physical and emotional functioning, pain, anxiety and depression, and perceived health, at post-intervention and at 12 months after intervention, and *(Study IV)*

- to evaluate the comparative effectiveness between IMPR of short, moderate and long duration on these outcomes. *(Study IV)*

4 METHODS

4.1 STUDY DESIGNS

Four studies are included in the thesis. *Study I* is a systematic review and meta-analysis of original international reports with a prospective study design. *Study II* is a methodological study, on inter-rater agreement. *Studies III* and *IV* are of prospective, large-scale multicenter, observational designs; *Study III* focuses on prognosis and *Study IV* on IMPR effectiveness. Study design, participants, data sources, and analyses are summarized in Table I.

Table I. Overview of study design, participants, data sources and analyses of the included studies.

	Study I	Study II	Study III	Study IV
Design	Systematic Review with meta-analysis	Inter-rater-agreement (reliability)	Pragmatic longitudinal multicentre register-based study.	Pragmatic longitudinal multicentre register-based *controlled* study.
Material/ Participants	25 original research reports, published 1983-2016 n= 9436	2 independent raters 43 research reports	38 specialist clinics n=2876, complete cases: baseline and 12 month follow-up after IMPR	15 specialist clinics n=2413, baseline, post and 12 month follow-up after IMPR
Data sources	Medline, Embase, PsychINFO, CINAHL, Web of Science, Cochrane CENTRAL		Swedish Quality Registry for Pain Rehabilitation	
Analysis	Narrative and Quantitative meta-synthesis Grading of Level of Evidence	Cohen's quadratic weighted kappa coefficient	PCA for dependent variables Logistic regression	GEE- within/between group analyses Proportion change one MCID

GEE: Generalized estimating equation, IMPR: Interdisciplinary multimodal pain rehabilitation, MCID: Minimal clinical important difference, PCA: Principal component analysis

4.2 ETHICS APPROVAL

Studies I and II. Analyses were based on already published data included in systematic reviews of original studies *(Study I)* and resulting data from Risk of Bias rating performed by the authors themselves as part of processing the systematic review material *(Study II).* There were therefore no personal data or data of sensitive nature generated in this review project, hence no study-specific ethical permission was required for *Studies I-II*. The protocol is registered in the International Prospective Register of Systematic Reviews (PROSPERO) id:

CRD42016025339, and a study protocol presenting the design and planned evidence work was published in advance of the main study (79).

Studies III and IV. These studies constitute part of a Swedish national project involving original nationwide registry data from the SQRP. The project has been granted ethical approval (Etikprövningsnämnden Stockholm, dnr: 2013/1842-31/2) (covering *Studies III-IV*) and the protocol is registered at Clinical Trials.gov (id: NCT02248363). All data for the purposes of the present project was made anonymous by statisticians at SQRP before exporting it, hence no identifiable personal information was included when data is received by the project group. Furthermore, analysis and presentation of the data are only conducted at group level and hence not possible to identify specific individuals by such analysis/presentation.

4.3 PARTICIPANTS AND SETTINGS

The study population in this thesis consists of working-age adults with chronic musculoskeletal pain for a duration of over three months, not emanating from malignancies or systemic diseases, e.g. back or neck pain, generalized pain syndromes, fibromyalgia, and general widespread pain, who have taken part in IMPR. Table II shows countries from which patients were recruited, and the patient characteristics age, sex, and pain duration for *Studies I – IV*.

Table II: Country, sex, age, mean pain duration (SD or range) of the included patients in the thesis (*Studies I -IV*).

	Study I	Study II	Study III	Study IV
Geographical origins/country	Europe (19), North America (5), New Zealand (1)	N.A.**	Sweden	Sweden
Sex, women%	60%*	N.A.	77%	80%
Age, years (SD)	43 (10.4)	N.A.	43.5 (10.7)	41.3 (10.7)
Pain duration, mean (range)	3 months- > 10 years	N.A.	106 months 66 median	79 months

SD: Standard deviation *For *Study I*, the proportion of females was calculated from the information available from all included studies, and for pain duration, the mean and range of the included studies was calculated. **Study II is based only on material derived from the Risk of Bias ratings of included articles, not the reported patient data.

As relevant for *Studies III – IV*, Figure 2 illustrates the geographical distribution of specialist IMRP clinics in Sweden in relation to population density (all 38 clinics marked were included in *Study III*, while 15 of these clinics were included in *Study IV*).

Figure 2. IMPR clinics in Sweden, linked to SQRP and population density.

4.4 INTERVENTION

In *Studies I-II*, our inclusion criteria was participation in any multidisciplinary, interdisciplinary or multimodal rehabilitation, with a biopsychosocial treatment approach, coordinated by ≥3 different health professionals where a shared collaboration was explicitly expressed. Health professionals involved were commonly a physician, a physiotherapist and a psychologist, with a social worker or occupational therapist on occasion. IMPR (in these studies referred to as MDR) could be of any duration/intensity and rehabilitation approach, in either inpatient or outpatient settings. IMPR-program duration commonly lasted between two to eight weeks, with nearly half of the studies reporting an IMPR duration of four to eight weeks. The remaining studies reported a duration exceeding eight weeks. Some programs included two phases, with a follow-up period with prolonged rehabilitation time. IMPR was either delivered as full-time treatment over a period of a couple of weeks, or part-time treatment over a couple of months.

In Sweden, the definition for IMPR is basically identical to the criteria given for *Studies I-II* above. Specifically, Swedish national medical guidelines for IMPR for chronic pain define it as treatment with an interdisciplinary team specializing in pain management, consisting of at least three professions; a physician specializing in rehabilitation medicine, a physiotherapist, a social worker, and a psychologist, plus an occupational therapist if needed (23). Commonly, IMPR is scheduled as a group-based program, with groups usually consisting of six to eight participants. It is emphasized that within these groups, each patient has an individual plan, schedule, and contacts, according to the patient's own goal setting. IMPR programs typically last about 10 weeks, but range between 4-18 weeks, with treatment delivered 1-5 days/week, in half or full days, for 8-40 hours/week. Some programs run continuously with face-to-face meetings, while others are delivered with an initiation phase, a self-management phase, and a completion or booster phase.

4.5 OUTCOMES

In order to measure general functioning and the impact of chronic pain on more general life concepts the Short Form Health Survey (SF-36) was used as a main outcome measure. The SF-36 is a self-assessed multi-purpose short-form health survey with 36 questions reflecting health-related quality of life. The 36 included questions yield an 8-scale profile of functional health and wellbeing scores, which can be incorporated into the psychometrically-based physical and mental health summary measures; SF-36 Physical Component Summary (SF-36 PCS) and SF-36 Mental Component Summary (SF-36 MCS). These two summary measures of SF-36, both ranging from 0-100, were used as dependent measures of physical and mental functioning. The SF-36 has been proven useful in surveys of general and specific populations, for comparing the relative burden of diseases and in differentiating the health benefits produced by a wide range of different treatments (24). The outcomes in this thesis relate mainly to general physical and psychological functioning, and expand though more proximal to more distal disease impact concepts. For instance, the SF- 36, was used in its more proximal form in *Study I*, with the SF-36 subscale Physical functioning (PF), reflecting functioning from a more delimited level while in *Study III* the SF -36 Physical Component Summary was used as a reflection of general physical health and functioning, the same applied for the equivalent SF-36 Mental Component Summary. If looking at these two component summaries in tandem they can apply as measures of more distal general concepts such as HRQoL. By using the summary scores of SF-35 PCS and MCS, our outcomes are thus closer related to concepts of general functioning. In *Study IV* we also evaluated an array of measures, covering a bio-psycho-social spectra, and at the same time core, to proximal-distal disease impact concepts. All of these constitute the mandatory measures in SQRP.

For a graphical illustration according to Schematic diagram depicting the outcome measures used in Studies I and III-IV – and their main position on subdomains of proximal-distal concepts of functioning see Figure 3, and for further details, see below:

The patient's reported pain intensity for the last week (7 days) was quantified with the Numeric Rating Scale (NRS), administered by asking the patient to estimate his or her pain on a scale of 0 to 10, with 0 representing no pain and 10 representing the worst possible pain (25)

Perceived health status was assessed using the EuroQol-5 dimensions (EQ-5D). This is a self-assessed standardized measure that provides a simple, generic measure of health for clinical and economic appraisal. It is applicable to a wide range of health conditions and treatments and it provides a single index value for health status (where a score of 1 represents good health and -0.564 "worse than death" that can be used in clinical and economic evaluations of health care as well as in population health surveys (26).

Further, and more specific to the experience of chronic pain, dimensions of pain intensity, emotional distress, cognitive and functional adaptation, and social support were assessed using the Multidimensional Pain Inventory (MPI) (27). In accordance with the procedure described by McKillop et al., two summary scores were used for the presentation of domain impairment and social support, MPI Impairment (0-6, where high scores represent high perceived impairment) and MPI Social support (0-6, where high scores represent high social support) (28).

Finally, symptoms of anxiety and depression in patients with somatic diseases were assessed using the Hospital Anxiety and Depression Scale (HADS), which is a widely used patient self-rated scale with 14 questions; 7 addressing "anxiety" and 7 addressing "depression" (subscales; HADS-A – anxiety and HADS-D – depression) that each ranges from 0-21 (in total 0-42) (29).

Study II did not include any patient reported measures, outcome was instead the degree of inter-rater agreement, as assessed with Cohen's weighted kappa. For more specific study operationalization, see details reported in sections below.

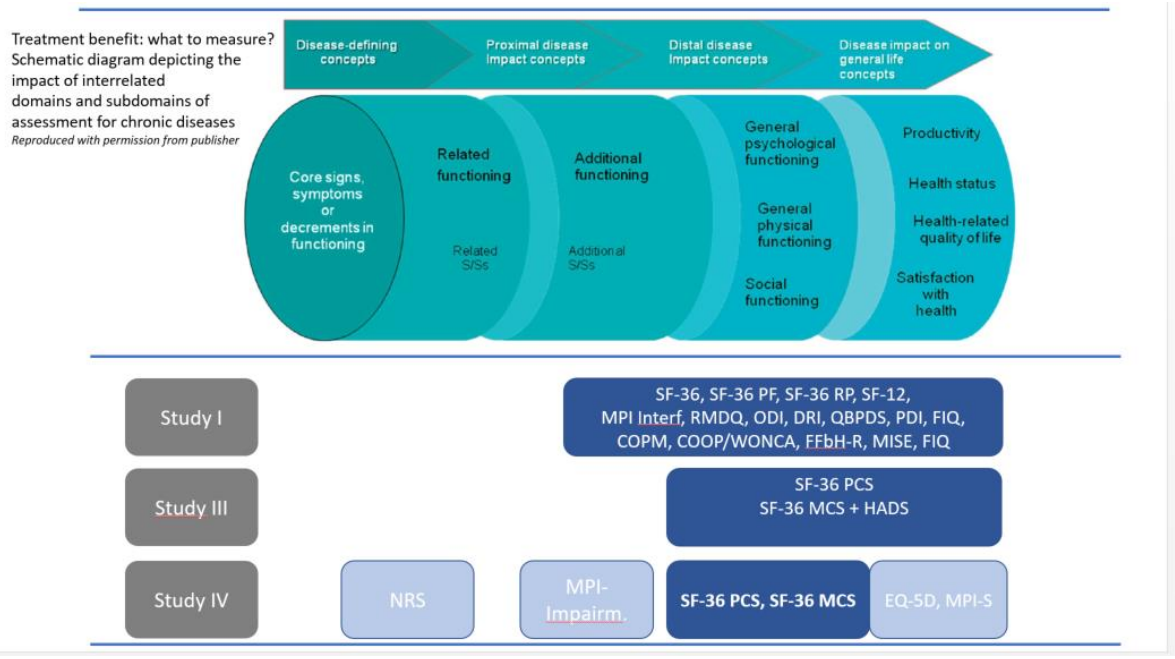


Figure 3. Presentation of outcome measures used in Study I and III-IV - and their main position on subdomains of proximal- distal concepts of functioning (56). Dark blue represents main outcomes, light blue represents secondary outcomes

4.6 STUDY I

4.6.1 Data sources and Material

For the purpose of identifying original studies of relevance, electronic searches were performed in six databases of major relevance for the topic: MEDLINE (via Ovid), EMBASE (via Elsevier), PsycINFO (via Ovid), CINAHL (via EBSCO), Web of Science (via Thomson Reuters), and the Cochrane Central Register of Controlled Trials (CENTRAL). Searches covered the publication period January 1980 to April 2017.

A total of 25 original studies published between 1983 and 2016 were included in the study, all of which investigated and reported data on prognostic factors for physical functioning after IMPR. Included studies were conducted in Europe (19), North America (5), and New Zealand (1), with a total study population of 9436 participants taking part in IMPR.

4.6.2 Procedures

Study I constitutes the first part in a project entitled "Predictors of multidisciplinary rehabilitation outcomes in patients with chronic musculoskeletal pain", and focused on physical function – one of four targeted outcomes. For further details please refer to our published study protocol (79). The systematic literature review and meta-analysis was conducted according to the following four main steps, as recommended (80, 81).

1) Study selection,
2) Assessment of study quality (QUIPS),
3) Narrative and quantitative synthesis,
4) Grading of the levels of evidence (GRADE).

4.6.2.1 *Study selection*

The study selection process was performed using the Covidence online systematic review software (82): screening of titles, screening of title/abstract, screening of full texts, and a final screening of study relevance, subsequently excluding studies revealing non-eligibility criteria. Figure 4, The Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) Flow Diagram (83), shows the study selection process and the number of papers included in the different stages of the systematic review.

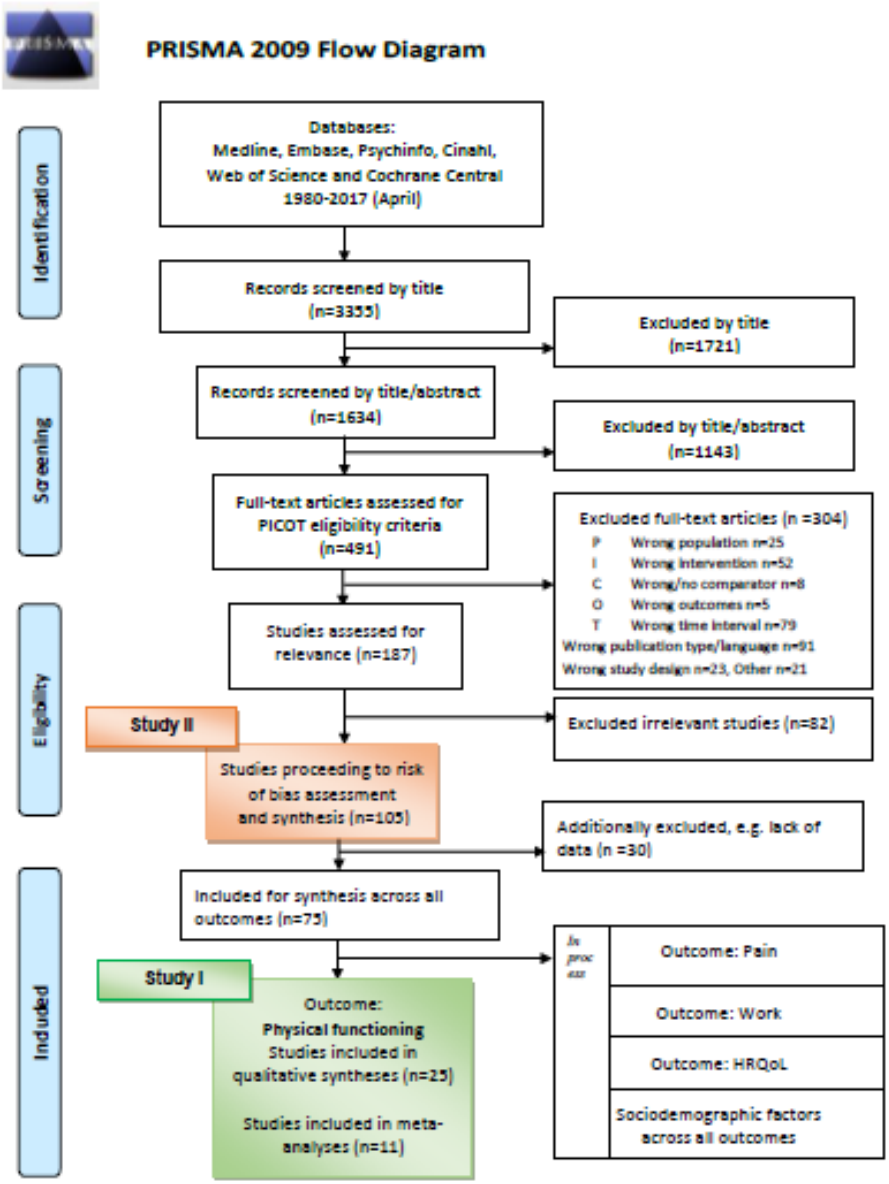


Figure 4. PRISMA Flow Diagram illustrating systematic review process, and Study I and Study II material. (84) PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
Tseli et al. *Clin J Pain* 2019 35(2):148-173, modified under the Creative Commons Attribution License 4.0.

A multidisciplinary review team, consisting of six members (two physicians specializing in rehabilitation medicine, one psychologist, and three physiotherapists), performed the screening and RoB assessments, independently and two by two. A randomization scheme was used to distribute studies in two evaluation trails, each with a set of independent reviewers. For more details please refer to study protocol (79).

4.6.2.2 *Study quality*

The studies that fulfilled the eligibility and relevance criteria were assessed using the Quality in Prognostic Studies (QUIPS) tool (69) to evaluate the Risk of Bias (RoB). All studies were rated on a three-point scale: "low", "moderate", or "high" RoB.

4.6.2.3 *Narrative and quantitative synthesis*

In the part of the study that focused on prognostic factors for physical functioning, papers with this specific outcome were selected (n= 25). Data from every study qualifying for inclusion in the present study were extracted into a coding file which provided the material for narrative and quantitative syntheses. In the narrative analysis, the results of each prognostic factor in relation to outcome physical functioning were scored with + or -, indicating the direction of a statistical significant association between the prognostic factor and the outcome. Those prognostic factors that did not reach statistical relevance were identified with a zero.

After the narrative synthesis, a more specific evaluation of statistical outcomes of prognostic data was carried out by performing a meta-analysis. Effect sizes from single prognostic factors were computed and transformed when necessary into a common index, odds ratio (OR), and 95% confidence interval (CI). The variance-weighted pooled ORs were then computed through a random effects meta-analysis using the software Review Manager (85). Heterogeneity was assessed, and funnel plots and sensitivity analyses were performed accordingly (86).

4.6.2.4 *Grading the evidence*

For the evidence synthesis, the analyzed prognostic factors were evaluated with the GRADE method, judging the confidence in the effect estimate of the prognostic factor on a four-point scale, including "very low", "low", "moderate", and "high" level of evidence (87, 88). In agreement with previous reporting on adaptation of GRADE in prognostic studies (88), the initial level of evidence was set at high quality and certainty the evidence was thereafter assessed based on the remaining eight criteria; RoB, imprecision, inconsistency, indirectness, and publication bias (decreasing evidence level) and large magnitudes of effect, dose-response gradients, and the lack of confounding (increasing evidence level).

4.6.3 Assessments

The dependent variable *Physical functioning* was assessed primarily through self-reports, for example using the Short Form 36-item Health Survey (SF-36), Roland-Morris Disability Questionnaire (RMDQ), or the Oswestry Disability Inventory (ODI). Included measures were generic or disease-specific, targeting functioning or, the opposite, i.e. disability. Included measures extended across proximal and distal impact concepts, see Figure 3 for a description of measures.

The independent variables: Measures of investigated potential prognostic factors were collated into relevant assessment domains; Pain-related factors, Physical function-related factors, and

Psychological factors. For each domain, the included measures were further partitioned into seven subdomains before synthesis, each representing a prognostic factor: Pain intensity, Pain duration, Self-reported physical function, Performance-based physical function, Emotional distress, Cognitive Behavioral 'Protective' factors (positive direction), and Cognitive Behavioral 'Risk' factors (negative direction). A detailed description of included measures within each prognostic factor is given in *Study I.*

4.7 STUDY II

4.7.1 Material

The 105 studies retrieved from the systematic search as described in the study protocol (79) and *Study I* formed the base of *Study II.* The studies that were evaluated for RoB by the two experienced raters independently during the systematic review process were analyzed for inter-rater reliability; this included a total of 43 papers, 10 of which pertained to the outcome Physical functioning *(Study I*) and an additional 34 pertaining to other outcomes, i.e. pain, work and quality of life.

4.7.2 Procedures

To assess the RoB in the studies selected for inclusion, a new tool, the Quality In Prognosis Studies (QUIPS) was used. QUIPS, developed for quality assessment in prognostic studies, consists of six domains; 1) Study participation, 2) Study attrition, 3) Prognostic factor measurement, 4) Outcome measurement, 5) Study confounding, and 6) Statistical analysis and reporting – and these are rated on a three-point scale with low, moderate, or high RoB. Each domain has 3-7 corresponding prompting items, to be answered on a four-point scale (yes, partial, no, unsure). These are intended to provide a basis for judgment on RoB for the domain in question (69). For every specific use of the instrument, a key list is required, which was created by the raters before the initiation of the quality assessment, and suggestions on how to create this were provided by the instrument developer through personal communication.

In order to become familiar with the quality assessment process, some studies were strategically selected for a common discussion on ratings. Thereafter, a first round of RoB assessment was performed independently (n=15) and the results were discussed in order to reach consensus. The remaining studies were then assessed, again independently, in a second round (n=28) and conflicts were resolved as they were previously.

After completion of the RoB assessment, evaluation of the inter-rater agreement was performed using Cohen's quadratic weighted kappa coefficient of both the total number of studies and for round 1 and 2 separately. As a final step, the authors discussed the results and elaborated on the functionality of the tool, highlighting the difficulties and strengths of the tool to aid future studies using QUIPS.

4.8 STUDY III

4.8.1 Data source and Material

The SQRP provided data, and the present study was based on 38 specialist care clinics across Sweden, including 2,876 patients who both participated in an IMPR program during the period 2012 - 2014 and took part in the 12-month follow-up. The geographical distribution of these clinics included all parts of Sweden, and was proportional to population. Further details regarding study procedures are given in *Study III.*

4.8.2 Dependent and independent variables

Dependent outcome variables were defined using Principal Component Analysis (PCA). Our PCA with orthogonal rotation emerged with two distinct components, relating to *Physical functioning* based on the SF-36 PCS solely, and *Emotional functioning* based on SF-36 MCS, HADS-A, and HADS-D respectively, hence reflecting a general perspective on functioning and HRQoL. To create binominal outcomes, we used the Minimal Clinical Important Difference (MCID) as cut-off, for a definition of improvement on our dependent outcomes. An increase of ≥3 on SF-36 PCS was chosen as a MCID for Physical functioning, and for *Emotional functioning*, the preconditions were an increase of ≥3 on SF-36 MCS and a decrease of ≥1.5 on either HADS-D or HADS-A.

Regarding independent variables, there were 18 baseline regressors (potential prognostic factors) used in the initial regression analyses. The variables represented socio-demographic, pain-related, and other multidimensional health descriptors, a detailed description of included independent variables is given in *Study III*. A theoretical and empirical reasoning (89) preceded the selection of these baseline regressors (89), which were then entered in univariate and multivariable analyses.

4.9 STUDY IV

4.9.1 Data source and Material

The SQRP again provided data, and *Study IV* was based on data from 15 specialist care clinics across Sweden. It included 2,413 patients who both participated in an IMPR program during the period 2012 - 2014 and were followed through post-intervention and 12-month follow-up. The reason for including 15 clinics was based on the intention to have reliable data for the assessment of comparative effectiveness of treatment duration. Here, as a complement to new and previously untested register data on the duration of IMPR, a crosscheck validation was carried out for data of relevance (representing 2012 - 2014), using independent, empirical information from clinic representatives. This yielded 15 clinical settings for which baseline

data and PROMs were extracted for analysis. The geographical distribution of these 15 clinics included all parts of Sweden, was again proportional to population, and the range of program duration was also equally dispersed. Further details regarding study procedures is given in *Study IV*.

4.9.2 Dependent and independent variables

Primary outcomes were *Physical* and *Emotional* functioning, based on the psychometrically-based physical and mental health summary of Short Form 36-item Health Survey; SF-36 PCS and SF-36 MCS, representing core domains of HRQoL (5). Secondary outcomes were Pain intensity (NRS), two summary scales derived from the Multidimensional Pain Inventory (MPI); perceived impairment and perceived social support, HADS, and Perceived health (EQ-5D). Effectiveness within and between groups was evaluated in two ways, estimating the marginal effects using Generalized estimating equation (GEE), and by calculating percentage attaining a MCID. The MCID was used as cut-off for a definition of improvement on our dependent outcomes. The MCID was also used in the opposite direction, reverting the same cut off as an indicator of possible deterioration on our dependent outcomes. Hence three possible outcome categories were identified on the presented MCID; improved, no change, and deteriorated. Further details on scale range, direction, and MCID-definitions of included outcome measures are presented in Table III.

Table III. Patient-reported outcome measures and MCID-definitions

	Subscale	Scale range	What is 'better'	MCID-levels	Reference
Primary outcome measure					
Short Form 36-item Health Survey (SF-36)	Physical Component Summary	0-100	↑	≥3	(90, 91)
	Mental Component Summary	0-100	↑	≥3	
Secondary outcome measures					
Numeric Rating Scale (NRS)		0-10	↓	≥ 2	(92, 93)
Multidimensional Pain Inventory (MPI) (94)	Impairment summary scale	0-6	↓	≥ 0.6	(95)
	Social support summary scale	0-6	↑	≥ 0.6	
Hospital Anxiety and Depression Scale	HADS Anxiety	0-21	↓	≥1.5	(96)
	HADS Depression	0-21	↓	≥1.5	
European Quality of Life instrument (EQ-5D)	EQ-5D index	-0,564-1	↑	≥ 0.1	(97)

Independent variable: IMPR duration (weeks). For the categorization of the variable, clinics were divided into three delimited and pragmatic groups of different program duration, intended to reflect current diversity of applied IMPR: short (4-9 weeks; 7 clinics), moderate (10 weeks; 3 clinics), and long (11-18 weeks; 5 clinics).

4.10 STATISTICAL ANALYSES

The statistical analyses used in this thesis are presented in Table IV. For details on statistics, see Supplementary Study I – IV. Table IV shows statistical analyses used in this thesis.

Table IV. Statistical analyses used in this thesis.

Statistics	Study I	Study II	Study III	Study IV
Descriptive				
Frequency (n), percentage (%)	x	x	x	x
Mean/median (SD/range)	x		x	x
Effect size calculation, and conversion to common index: Odds Ratio	x			
Inferential				
Cohen's kappa	x	x		
Meta-analysis	x			
Funnel plots	x			
Pearson's correlation coefficient			x	x
Principal Component Analysis			x	
Minimal Clinical Important Difference			x	x
Logistic regression			x	
Generalized estimating equation (GEE)				x

4.10.1 Study I

The strength of the relationship between a specific prognostic factor and corresponding outcome was quantified using the specific commonly-used procedures as descripted by Lipsey (98). In short, the statistical outcomes (effect sizes) from single prognostic factors were with the help of online calculators converted into a common index of variance-weighed ORs to permit pooling across studies (86, 99, 100). For each prognostic factor, the software Review Manager (85) was used to compute variance weighted pooled ORs and 95% confidence intervals (CI) in a random effects model. The generic inverse variance method was used, since this method permits a wide selection of data formats in the analyses (101). Measures of statistical heterogeneity were calculated (e.g. I^2), and funnel plots were used to assess potential publication bias (86, 99, 100).

4.10.2 Study II

The inter-rater agreement was investigated to determine the consistency between the two experienced senior raters using all domains from all papers (n=43) that were assessed by these raters. The raters judged the papers separately using the QUIPS three-graded RoB scores (”low”, ”moderate”, or ”high” RoB). When the two raters scored alike on a domain, this was considered as an “agreement”, while “maximal disagreement” was defined as one rater scoring low RoB and the other scoring high RoB. All other cases of disagreement were defined as partial disagreement. The percentage of the number of domains judged alike by the two raters divided by the total number of domains (n= 258; i.e. 43 papers × 6 domains) before reaching consensus was calculated as a rate of agreement across all papers. This was also calculated for separately for the first and second rounds (n= 90 and n= 168, respectively). An online kappa calculator was used to calculate Cohen’s kappa (κ) with quadratic weighting (102) as a measure of inter-rater reliability and the coefficient was interpreted as follows: κ= 0.00–0.20, no agreement; κ= 0.21–0.39, weak agreement; κ= 0.40–0.59, minimal agreement; κ= 0.60–0.79, moderate agreement; κ= 0.80–0.90, strong agreement; and κ> 0.90, almost perfect agreement. Moreover, we tested if there were any differences in overall agreement between the raters, between the different rounds, and between the papers with high and low RoB by using an online test of proportion calculator (103).

4.10.3 Study III

Binary logistic regression was initially used to quantify the univariate associations between baseline variables (regressors) and outcomes, and an inclusion criterion of $p \leq 0.2$ was preferred for further analyses in multiple regression models. Regressors were examined for potential multicollinearity bias.

Using a top-down approach, the final multiple regression model was identified in a 2-step procedure. Firstly, stepwise backwards elimination was used to eliminate variables based on the highest p-value, until only variables significant at $p \leq 0.2$ remained. Secondly, in accordance with the purposeful selection process described by Hosmer, Lemeshow, and Sturdivant (34), variables already eliminated in the univariate analyses were included one by one in the multiple regression model and retained if they were significant at $p < 0.05$, yielding a preliminary final model with all variables significant at $p \leq 0.2$. Baseline regressors significant at $p < 0.05$ in the final model were recognized as important prognostic factors.

4.10.4 Study IV

Regarding treatment effectiveness on duration, a repeated GEE with robust standard errors was used to compute the average within-subjects effect for the complete sample from baseline to rehabilitation end, and to one-year follow-up (104). Standardization was used to compute the marginal effects between the duration groups, based on a general linear model with robust standard errors (105). In our between-group analyses, IMPR duration (*short* vs. *moderate* vs. *long*) was used as a between-subject factor. Treatment days/weeks and baseline level in all defined outcomes were adjusted for in the GEE modeling. GEE handle missing data by

including all available data in the estimations. Alpha levels were set to $p < 0.01$ for all analysis stages. Our sample size of 2413 patients entailed high test-power for the analyses. Therefore, in addition to tests of statistical significance, all outcomes were dichotomised into patients who attained a MCID and patients who did not (106), in order to aid clinical interpretation.

5 RESULTS

5.1 STUDY I

A summary of findings, and overall quality as assessed with GRADE, is shown in Table VI. Results from the evidence synthesis on prognostic factors for physical functioning at ≥ 6 months is presented for the three prognostic factor domains; Pain, Physical, and Psychological. *Pain related factors:* Pain intensity and pain duration were not associated with outcome. Here, the level of evidence was moderate due to inconsistency of the results. *Physical factors:* High self-reported physical functioning was associated with a positive outcome. The level of evidence was low, due to inconsistency of the results and study limitations, i.e. risk of bias. Performance-based physical factors, analyzed with narrative synthesis only, were not associated with outcome and there was no evidence of prognostic value. *Psychological factors:* Low levels of emotional distress [(A)], high levels of cognitive behavioral protective- factors [(B)], and low levels of cognitive behavioral risk-factors were associated with a positive outcome [(C)], see Figure 5. The level of evidence was moderate for all three psychological factor-groups.

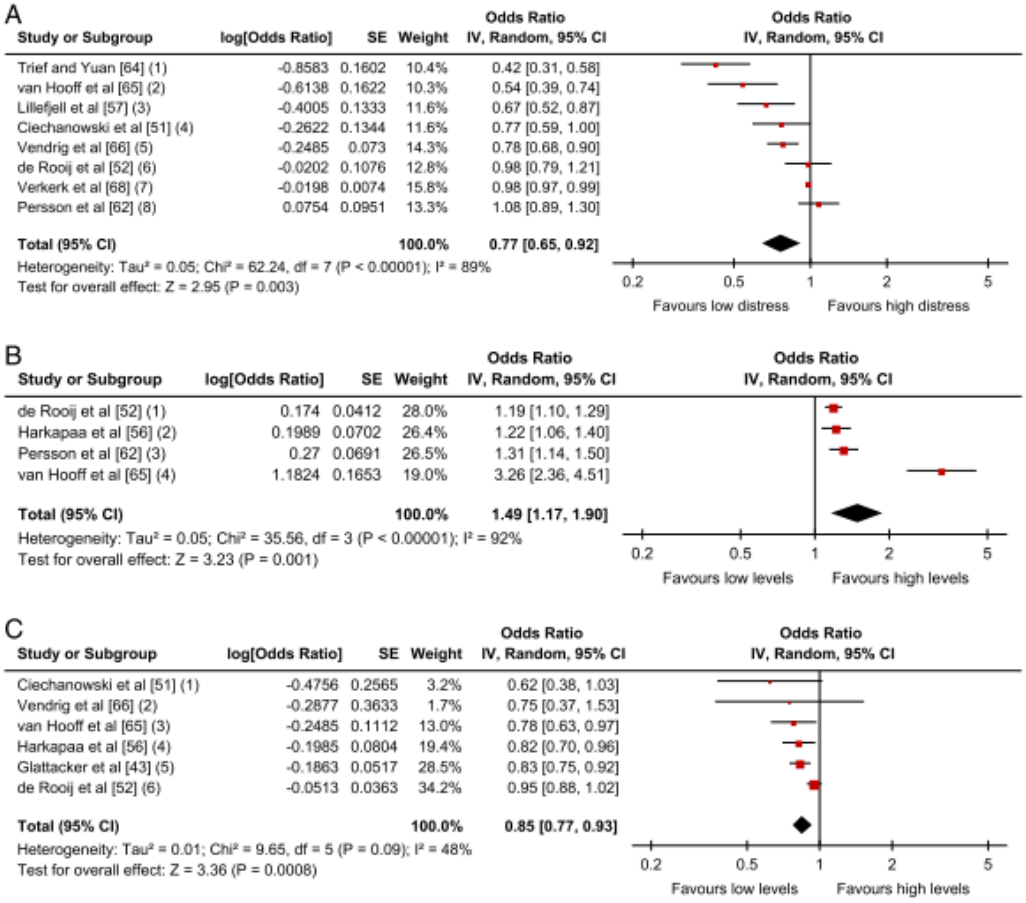


Figure 5. Forest plots for Psychological factors and their association with a positive outcome: A) emotional distress, B) cognitive behavioral risk factors, and C) cognitive behavioral protective factors (84).

Table VI. Summary of findings and GRADE - level of evidence (84).

TABLE 7. Summary of Findings and Overall Quality as Assessed With GRADE

		All Studies	Studies Included in the Meta-Analysis		GRADE Factors							
Domain	Potential Prognostic Factor	Total Number of Participants (No. Studies)	Total Number of Participants (No. Studies)	Estimated Effect Size (95% Confidence Interval)*	Phase	Study Limitations	Inconsistency	Indirectness	Imprecision	Publication Bias	Moderate/ Large Effect Size	Dose Effect	Overall Quality (Level of Evidence)
Pain	Pain level	8191 (16)	2676 (5)	OR, 0.84 (0.65-1.07)	+++ +	0	—	0	0	0	0	0	Moderate quality (+++)
	Pain duration	3800 (8)	2978 (5)	OR, 0.97 (0.93-1.00)	+++ +	0	—	0	0	0	0	0	Moderate quality (+++)
Physical	Performance-based function	783 (2)	NA (0)	NA	+++	—	—	0	0	0	0	0	Very low quality (−)
	Self-reported function	4706 (14)	3444 (8)	**OR, 1.07 (1.02-1.13)**	+++ +	—	—	0	0	0	0	0	Low quality (++)
Psychological	Emotional functioning	4358 (15)	3483 (8)	**OR, 0.77 (0.65-0.92)**	+++ +	—	0	0	0	0	0	0	Moderate quality (+++)
	Cognitive and behavioral protective factors	2288 (9)	1392 (4)	**OR, 1.49 (1.17-1.90)**	+++ +	0	0	0	0	—	0	0	Moderate quality (+++)
	Cognitive and behavioral risk factors	4068 (11)	1173 (6)	**OR, 0.85 (0.77-0.93)**	+++ +	—	0	0	0	0	0	0	Moderate quality (+++)

Significant estimates in bold style.
GRADE indicates Grading of Recommendations Assessment, Development and Evaluation; NA, not available; OR, odds ratio.

Tseli et al. *Clin J Pain* 2019 35(2):148-173. Reprinted under the Creative Commons Attribution License 4.0 (CCBY).

5.2 STUDY II

The inter-rater agreement, when assessed for all domains and all included papers, was interpreted as "weak", κ = 0.475 (95% CI = 0.4–0.6). Round 1 emerged with "Minimal agreement" κ = 0.323 (95% CI = 0.1–0.5), and Round 2 emerged with "weak agreement" κ = 0.536 (95% CI = 0.4–0.7), indicating somewhat improved agreement after the conflicts/ consensus discussion following Round 1. Table VII shows inter-rater agreement in detail, for all rated papers and for the two rounds separately.

Table VII. Inter-rater agreement for all rated papers and for the two rounds separately (107).

Table 2 Interrater agreement for all papers and for the two rounds separately

	All papers (43 papers)		Round 1 (15 papers)		Round 2 (28 papers)	
	κ^1	95% CI[2]	κ^1	95% CI[2]	κ^1	95% CI[2]
1. Study participation	0.356	0.063–0.669	− 0.023	− 0.041–0.594	0.475	0.100–0.851
2. Study attrition	0.647	0.353–0.771	0.400	0.000–0.835	0.714	0.426–1.000
3. Prognostic factor measurement	0.384	0.044–0.723	0.852	0.544–1.00	0.028	− 0.453–0.612
4. Outcome measurement	0.358	0.072–0.645	0.211	0.00–0.660	0.443	0–095-0.792
5. Study confounding	0.526	0.269–0.784	0.561	0.00–1.00	0.481	0.186–0.776
6. Statistical analysis and reporting	0.364	0.077–0.651	− 0.070	0.321–0.486	0.533	0.176–0.627
Overall	0.475	0.358–0.601	0.323	0.129–0.517	0.536	0.390–0.682

[1]Quadratic weighted kappa (κ)
[2]95% Confidence Interval (95% CI)

Grooten et al. *Diagnostic and Prognostic Research* 2019 3:5. Reprinted under the Creative Commons Attribution License 4.0 (CCBY).

The degree of inter-rater agreement varied between the 6 assessed domains; the Kappa calculations showed minimal agreement for domains 1 "study participation", 3 "prognostic factor measurement", 4 "outcome measurement", and 6 "statistical analysis and reporting". They showed weak agreement for domain 5 "study confounding" and moderate agreement for domain 2 "study attrition".

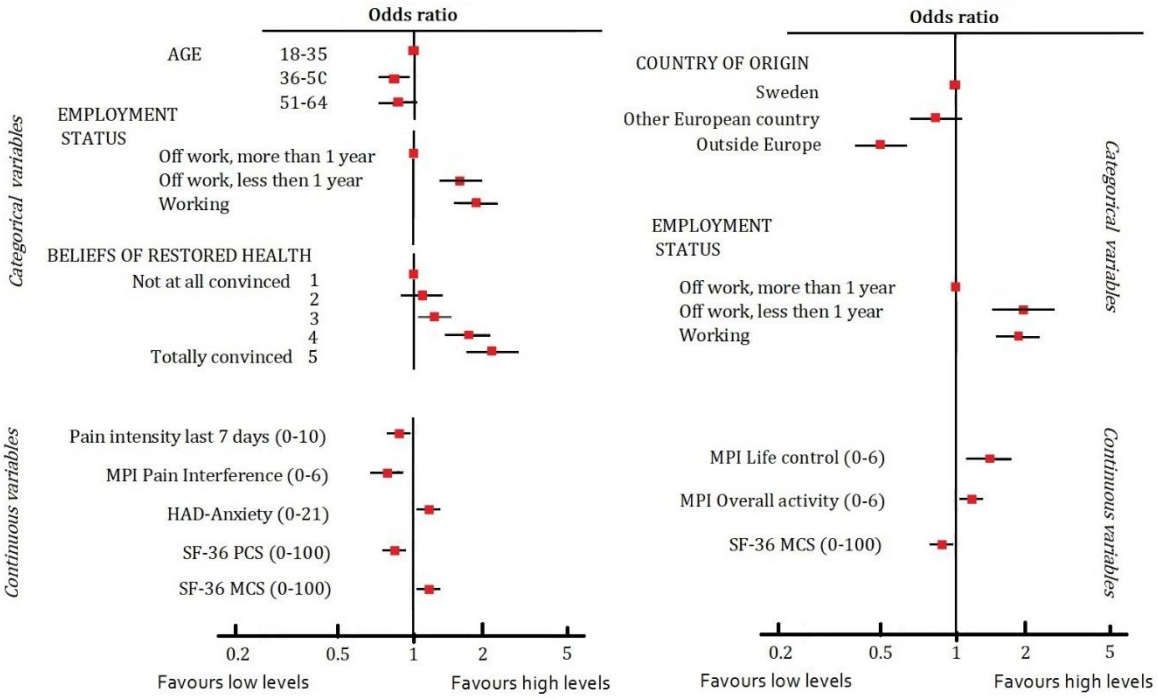
Elaboration of the results and suggestions for improving the inter-rater agreement and functionality are presented in a BMC Blog Network (108) accessed by the following link: http://blogs.biomedcentral.com/on-medicine/2019/04/18/assessing-study-quality-bias-prognosis-research-rehabilitation-chronic-pain-using-quips-tool/

5.3 STUDY III

Multivariable logistic regression analyses revealed the significant ($p \leq 0.05$) prognostic factors for improved *Physical and Emotional functioning* at 1-year follow-up after completion of an IMPR program – results are presented in text below and illustrated as forest plots in Figure 6.

Physical functioning: Employment status and one's beliefs in restored health were significant prognostic factors for improvement. Higher emotional functioning was also related, as were five additional variables: younger age, lower initial levels of pain intensity (NRS), pain-related interference in everyday life (MPI Pain interference), a lower level of physical functioning (SF-36 PCS), and higher initial levels of anxiety (HAD- A). In the final model, the overall percentage of correctly classified cases was 64% and the explained variance 15%, Nagelkerke R Square, 0.151 (Cox & Snell R Square 0.114). Diagnostic tests of the final model indicated acceptable discrimination (AUC of 0.70, 95% CI: 0.68-0.71).

Emotional functioning: Employment status proved again to be a prognostic factor for improvement. A lower level of emotional health (SF-36 MCS) emerged as related, as were three additional variables: a European origin, higher overall activity levels (MPI Overall activity), and higher sense of life control (MPI Life control). In this final model, the overall percentage of correctly classified cases was 67% and the explained variance 21%, Nagelkerke R Square, 0.210 (Cox & Snell R Square 0.157). Diagnostic tests of the final model indicated fair discrimination (AUC of 0.73 95%CI: 0.71- 0.75).



Physical functioning Emotional functioning

Figure 6. Prognostic factors for Physical and Emotional functioning one year after IMPR. Forest plot showing baseline socio-demographic, pain- and health-related factors, and associations with positive outcome, at $p \leq 0.05$. Categorical variables are shown with ORs for every category in comparison to reference category, and continuous variables are shown as "mean" OR for one step increase/decrease one range of specific measure (0-10, 0- 2I, 0-100). For SF- 36 PCS, SF-36 MCS, MPI Life control, and MPI Overall activity high scores = better health, and for Pain intensity, MPI Pain interference, and HADS-Anxiety high scores = worse health.

5.4 STUDY IV

In all, 2413 participants were referred to IMPR and completed an IMPR program at one of the 15 specialist clinics. Of these, a total of 1889 (78%) completed the post-intervention follow-up, and 1536 (64%) provided data at the 12-month follow-up. Analyses on baseline data revealed no clinically important differences between participants that provided data at the 12-month follow-up vs. those missing. Personal characteristics for the three groups were similar.

Effectiveness of IMPR – within-group analyses

IMPR consistently proved to be clinically effective on our main outcomes; physical and mental health. Patients improved significantly post intervention ($p < 0.01$) and the effects were sustained at the 12-month follow-up. This result was also evident for our secondary outcomes; pain intensity (NRS), Impairment and Social support (MPI), Anxiety and Depression (HADS), and Perceived health (EQ-5D). Figure 7 shows means and 95% CI, respectively, of the course in physical and mental health (SF-36 PCS and SF-36 MCS) from baseline to 12-month follow-up.

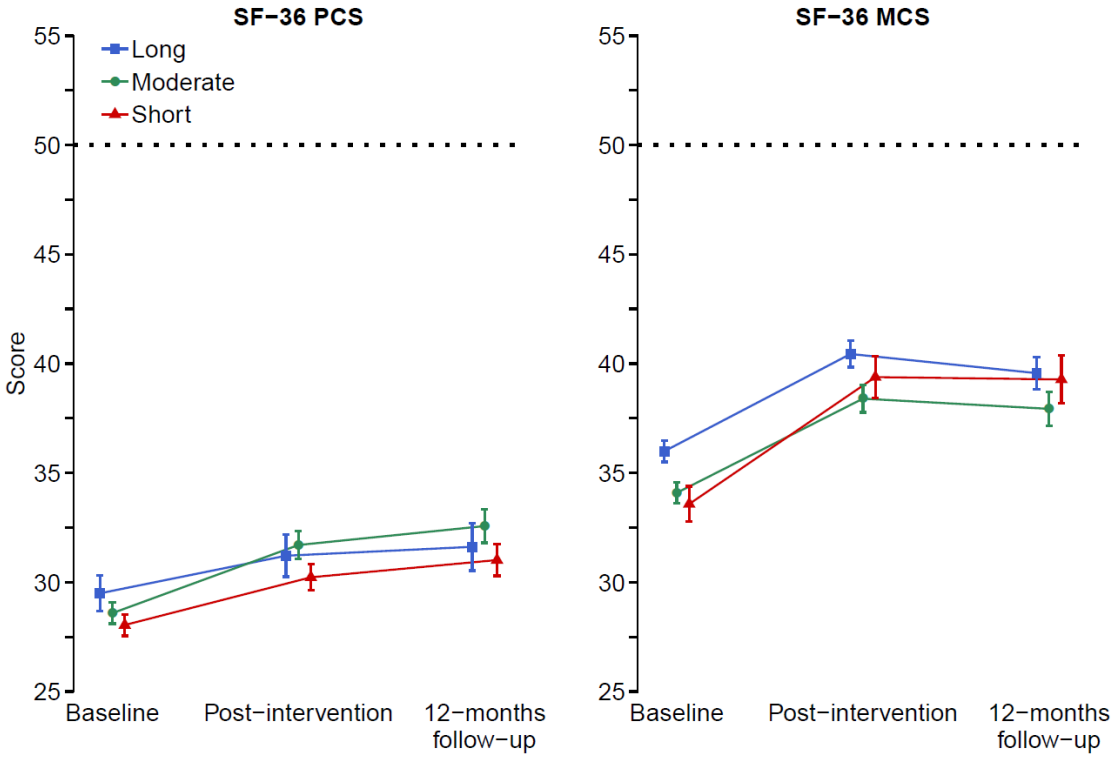


Figure 7. Change in HRQoL from baseline to post intervention and 12-month follow-up, measured using SF-36 Physical Component Summary (PCS) and Mental Component Summary (MCS). High values indicate better status. In the model, adjustment was made for baseline status and treatment days/week. Note: graphs show baseline values as they were, while statistical analyses were controlled for differences at baseline. The dotted line indicates the normal value for the general population

Effectiveness of IMPR – between-group analyses

There were no overall between-group effects; results showed that all groups, regardless of duration, had significant improvements in all outcomes, which were evident both post-intervention and at the 12-month follow-up. It was notable, however, that there was a statistically significant ($p = < 0.01$), but clinically negligible, greater 12-month effect on physical health in the moderate duration program (SF-36 PCS=0.009) than there was in the short duration program.

Despite the average effect being good, it emerged from our GEE-modeling that there was a noticeable proportion of patients who clinically deteriorated according to minimal clinical important difference (MCID) during the follow-up period. Here, regarding physical health, 42% improved significantly post-intervention according to MCID, however, 23% deteriorated according to MCID. At 12-month follow-up, this was 45% and 22% respectively. Regarding mental health, the corresponding MCID for improvement/deterioration was 52%/26% post-intervention and 50%/28% at the 12-month follow-up.

Figure 8 shows the proportion of patients attaining a change equivalent to a MCID at post-intervention and 12-month follow-up, on primary and secondary outcomes.

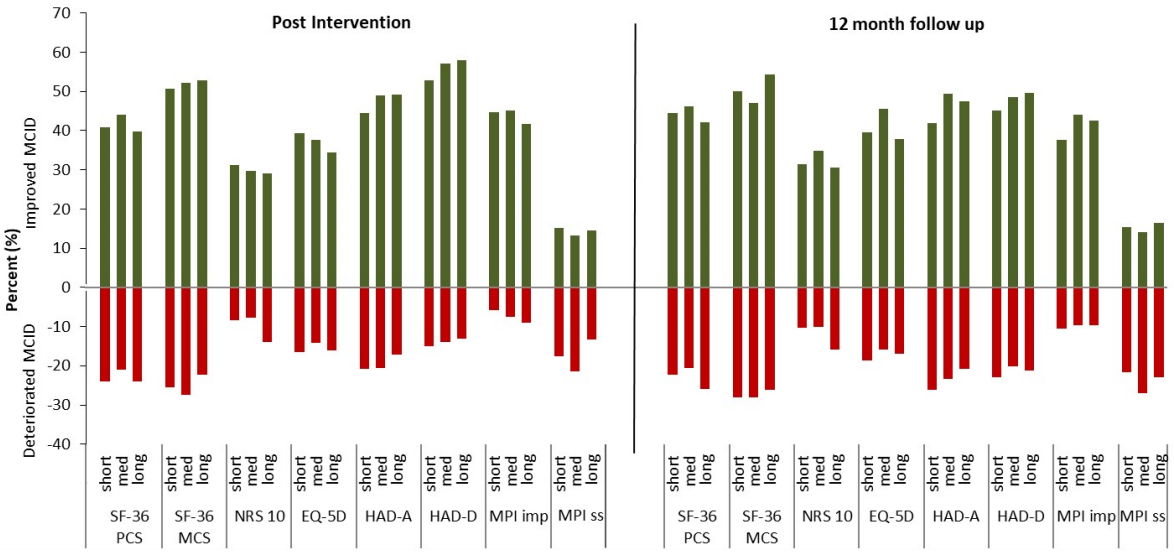


Figure 8. Minimal Clinically Important Difference (MCID) for primary and secondary outcomes at post-intervention and 12-month follow-up. The positive bars (green) show the proportion of patients who improved. while the negative bars (red) show the proportion of patients who deteriorated.

6 DISCUSSION

6.1 MAIN FINDINGS

The research presented in this thesis focuses on building evidence to tackle the contemporary challenges associated with rehabilitation in this major patient group of people suffering from chronic pain conditions. The evidence synthesis of published original studies showed, with a low to moderate level of evidence, that patient characteristics relating to better initial physical and emotional functioning, with less negative and more positive coping strategies, positively influenced prognosis on long-term physical functioning. Pain level and duration of pain were not prognostic for the outcome, with a moderate level of evidence, indicating that life impact concepts of pain are more relevant than pain itself for the understanding of prognosticating future functioning.

In the process of evidence building, RoB is certainly an important quality-assurance step to take. Our inter-rater agreement of QUIPS was initially weak and required a learning phase, but it improved somewhat after a learning phase was incorporated. Some domains and prompting items were more difficult to rate, and suggestions on how to improve RoB assessment in prognostic research in the field of chronic pain were provided.

Primary data on prognosis showed that patients who were working, or who had had only a short period of time off work, had a better prognosis for both main outcomes, *physical functioning* and *emotional functioning*. In line with results from the evidence synthesis, positive beliefs in recovery and 'good' initial emotional health were important for physical functioning. In contrast, a poor initial physical functioning indicated a better prognosis. Likewise, poor initial emotional health indicated a better prognosis for emotional functioning, but a strong sense of life control (i.e. less life impact) was also of importance. Hence, some baseline factors go in a 'better in better out' direction while others are 'the worse the better', presenting a complex prognostic picture for the complete understanding of good clinical follow-up on functioning and HRQoL. Regarding IMPR effectiveness, IMPR consistently proved to be clinically effective on our main outcomes as well as for our secondary outcomes. On the contrary, regarding the study hypothesis that a longer duration IMPR program would be superior to a shorter one, IMPR proved equally effective irrespective of program duration.

6.2 PROGNOSTIC FACTORS FOR A POSITIVE OUTCOME AFTER IMPR

The results from *Studies I* and *III* provide a basis for some important reflections. *Study I* investigated physical functioning based on PROMs only while *Study III* investigated both physical and emotional functioning, with the added value of incorporating socio-demographic factors in the prediction analyses as well. Common to both outcomes in the primary data was the importance of the *employment* variable, which showed that work was consistently a

protective factor for improvement. There seemed to be a linear trend, with a less positive effect as time off from work increased. Efforts to avoid delay therefore seem to be of importance for a better general prognosis. These results are further corroborated by other studies, also showing the importance of the work connection with HRQoL (109, 110).

A recent study, based on registry data from SQRP and linked with the Swedish Social Insurance Agency, investigated patterns of sick leave one year prior to IMPR and found an increase in sick leave prior to participating in IMPR, consisting of a larger percentage of full-time and partial sick leave, which later decreased again in the two years after rehabilitation. Rivano Fischer et al. emphasized the importance of further research into sick leave patterns before and after IMPR, in conjunction with PROM (111). Their study also showed that a longer period of sick leave before IMPR was found to be associated with lower rates of return to work (RTW), while those not on sick leave prior to IMPR had higher RTW. Further research into the prognostic value of work (and the time-period of sick leave prior to rehabilitation) is required. This kind of research, however, is difficult to perform and very nation and time specific, since policy alterations on sick leave, for instance, can change the meaning of the prognostic factor without a person having changed in other ways. Our cohort was taken from a period after the major revision in social insurance regulations in 2009. Also, further research into what it actually is in the 'Work-factor' in this specific population that constitutes its protective part is needed. One must also take into consideration that long-term sick leave might be a result of unsuitable work due to a patient's physical and mental condition that could prevent them from being able to return to working life in general.

The results from both the systematic review and meta-analyses (*Study I*) and primary data (*Study III*) showed that *physical functioning* was important for prognosis. When it comes to the more performance-based or objective prognostic factors there seems to be a lack of studies available using other data collection methods than PROMs for physical functioning. In our systematic review (*Study I*), we could not pool the estimates due to limited data, but our narrative synthesis showed no association for performance-based physical function, which was in line with van der Hulst et al. As indicated, however, when comparing associations on self-assessed physical function, we found, similar to van der Hulst et al., that these PROMs were of value in both our data from the systematic review/meta-analysis and primary data (59). Concerning the direction of this association, both in our narrative analyses and in van der Hulst et al., both high *and* low physical functioning could predict a positive outcome, leading to somewhat inconsistent results. Taken together, a similar discrepancy was found between *Studies I* and *III*, with the meta-analysis, in *Study I* indicating a positive association between high baseline functioning and improvement while the regression model in *Study III* showed a negative association. The reason for this is presently unclear. Except for initial physical functioning as a prognostic factor, however, the baseline prerequisites for an improved *physical functioning* seem to illustrate a good "better in better out" relationship, and indicate the importance of healthcare professionals working with behavior change in relation to exercise, activities, and increasing function. To improve, one needs to be in 'pretty good shape' mentally and have access to adaptive coping mechanisms. When it comes to the outcome *mental*

functioning, results from *Study III* indicated that people in worse shape are more likely to improve illustrating "the worse, the better". This is in line with findings in other recent studies (112, 113). One possible explanation could be that here it is the perceived support from the team and the environment that is the strengthening component. Further research is necessary to understand more of this somewhat complex clinical presentation.

For *Pain*, the prognostic value seems inconclusive as the results differ between *Studies I and III*. In the meta-analyses, the initial pain level and pain duration were not significantly associated with outcome, despite indicating the direction that low initial pain was better. In *Study III*, pain was somewhat associated with outcome, and a lower pain intensity was prognostic for improvement in physical functioning. Both studies indicated the same direction of the association, but this was not confirmed by the pooling of data. Perhaps this was due to a power problem as in previous reviews (58, 59). *Study III*, based on primary data, is uniquely large-scale and well powered, and if these new data were to be aggregated with previous forest plots in *Study I*, the effect size and confidence intervals would probably change.

The results from the systematic review and meta-analysis (*Study I*) showed that *psychological factors* were of importance. Results showed that low levels of emotional distress predicted a better physical functioning in long-term follow-up. This was partly supported by our primary data study, where high SF-36 MCS was significantly associated. Although the effect size for SF-36 MCS in *Study III* was small, one must remember that the presented odds ratio represents its continuous scale from 0-100. Parallel to this finding, another factor, HADS-A, emerged as significant, however, indicating the opposite: high anxiety was associated with improvement in physical functioning. It is difficult to explain why this result emerged, but it is believed that it could be related to IMPR effectively targeting aspects of anxiety, however this needs further study.

Cognitive and behavioral 'coping' factors presented a joint picture; low levels of cognitive and behavioral risk factors, such as fear avoidance and pain catastrophizing, predicted a better physical functioning at long-term follow-up. These results seem reasonable, as these factors are known to hinder functioning, which is why modern IMPR targets this domain in many ways, from pain school, to graded activity, exposure etc. Parallel to this, high levels of cognitive behavioral protective aspects, such as optimism, self-efficacy, and sense of control, were associated with high physical functioning. In this respect, positive affect, resilience, and sustainability are concepts that have received increased attention where chronic pain research is concerned. It is believed that an increased focus on positive, psychological, protective, and resilience factors may provide an opening to re-instituting best available physical function. Sustainability can be defined as the "ability of a person to move towards long-term positive outcomes in life in the presence of adversity". An increased focus on targeting not only recovery from disability but also the promotion of an upward spiral of sustainability, providing a buffer for the downward spiral seen in Figure 9, has been advocated (114).

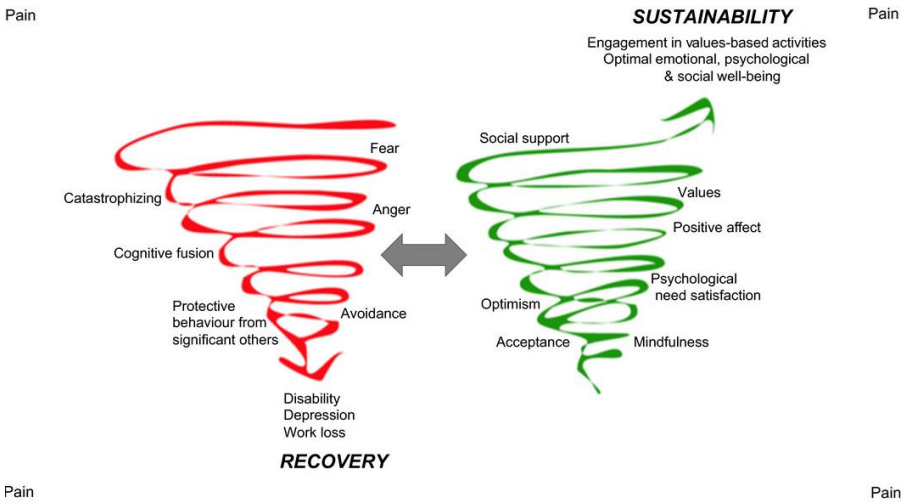


Figure 9. Recovery and sustainability in the field of chronic pain.
With permission from the publisher. Eur J Pain. 2017;21(8):1301-15. (114).

6.3 INTER-RATER AGREEMENT RELIABILITY OF QUIPS

The quality of the studies upon which evidence will be built is indeed of great importance (115) and, therefore, reliable instruments are necessary. Many other instruments use sum scores as an indication of total RoB, but in QUIPS the use of a sum score is not recommended as emphasis is put on scientific judgment and a sum score could give the appearance of false objectivity. It is important to understand that the instrument warrants a judgment to be made, and the final judgment is based on an informed discussion. Hence the assessment is not intended to be deterministic, and it also allows for some degree of subjectivity.

It was evident that it was not so easy to agree during the independent assessment, which was illustrated by the low kappa. On the other hand, the agreement increased over time in the second round. This highlights the presence of a learning phase. Also, it was evident that studies with good methodological quality were easier to agree upon. Sometimes judgment was also hindered by poorly reported data in studies, and here more variance in the interpretation was present. However, as the independent rating is just the first step of a scientific judgment, any discrepancies in judgment were revealed and dealt with. A poor inter-rater agreement in this stage is therefore not problematic for the final decision with regard to RoB and GRADE assessment – instead it is the prerequisite of a scientific discussion. We believe, however, that it is helpful to have knowledge on the inter-rater agreement analyses early on in the process to identify problematic areas as these may need more detailing, for which suggestions to clarification were given in *Study II*. Early inter-rater agreement analyses can also help reveal issues within (e.g. how many "nos" are required before downgrading a level on the RoB scale) or between particular domains, for instance we discovered that one methodological flaw in a study could be judged within two different domains, which resulted in low agreement until discovered.

It is difficult to compare *Study II* with previous studies on the inter-rater reliability of QUIPS due to a different context of prognostic research. Our study in the area of chronic pain was challenging due to the large variety of symptoms present, and the fact that we evaluate prognostic factors across multiple outcomes and other methodological difficulties such as low attrition in observational rehabilitation research.

6.4 EFFECTIVENESS OF IMPR

Following an IMPR program, patients improved significantly post intervention and the effects were sustained at the 12-month follow-up. Our findings are mostly consistent with several systematic reviews, which have concluded that IMPR generally is effective, but with small to moderate effects (21, 47, 50, 61, 63), and also in line with previous primary studies reporting long term improvements on pain, and quality of life (116, 117). Contrary to our findings and the general notion, a recent umbrella review assessed the strength of the evidence for effectiveness of IMPR and concluded that there is currently no robust evidence for the effectiveness of IMPR in any outcome, with the possible exception of decreased pain short term (118). One major contribution to this conclusion was the small number of studies included in the synthesis and the fact that many studies were susceptible to small sample sizes. Our results are based on large, real life empirical data, with wide national representativeness.

Importantly, although our results showed positive effects at a group level, data on our MCID indicated that around 20% of the patients deteriorated at least one MCID on the primary and secondary outcomes, and that is despite receiving IMPR. This is in line with recent findings from Vartiainen et al. (110), who also showed that while QoL was improved in nearly half of the patients, there were also approximately 30% who reported a clinically important deterioration at 12-month follow-up after rehabilitation. Until now, research results that show good (or no) overall group effects have rarely shown subgroups which, paradoxically, may tend to deteriorate, possibly because they often lack power to motivate such grouping. It is believed that the identification of subgroups has the potential to give important indications on how IMPR may be better tailored to each patient, and possibly also show which patient groups are in need of another type of rehabilitation. Further research presenting and further analyzing such bidirectional data is thus warranted to increase knowledge of possible explanatory characteristics.

6.5 EFFECT OF TREATMENT DURATION ON HRQOL

The motivation for emphasizing treatment duration as a between-group factor was that we hypothesized that it takes time for behavior change to be learned and absorbed into individuals' day-to-day routines (65). Contrary to our hypothesis, IMPR proved equally effective irrespective of program duration. Longer IMPR programs were not superior to shorter programs and although there was indeed a statistically significant between-group effect

favouring moderate vs. short programs, the difference was negligible and not detectable on MCID. Our findings are, however, in agreement with findings from systematic reviews that have examined the influence of treatment duration in the context of dosage (50, 118). Waterschoot et al. concluded that "on the basis of current literature it is unknown how many hours, months, or weeks are needed to achieve the best effects" (50). Using a meta-epidemiological approach, the meta-analysis by Dragioti et al. (118) examined the influence of treatment duration in the context of dosage, in patients with non-specific chronic low back pain, and they concluded likewise that treatment duration had no overall influence on the reported effects. However, in subgroup analyses, larger effect sizes for pain and disability emerged in favour of a long program with non-daily contact, while for quality of life, a shorter duration was favourable. In our study, however, the results were consistent across outcomes, with no between-group effects, or for SF-36 physical health, a statistical significant but clinically negligible effect favouring the moderate program. Both the above-mentioned systematic reviews based their evidence synthesis on available literature. Our results may add to theirs, by using large sample, pragmatic register-based primary data. We believe our study design and our results hold for what can be expected of a well-designed controlled study, by emulating, as far as possible, a register-based RCT.

6.6 METHODOLOGICAL CONSIDERATIONS

6.6.1 External validity Study I-IV

Regarding the generalization within the context of the intervention and the study population of interest, Table II shows that the patients included in the systematic review and meta-analysis *(Study I)* are corresponding well to the patients included in the national registries (*Studies III and IV*), in respect to gender, age, and pain duration. In *Studies I, III,* and *IV*, evidence was derived from longitudinal and/or pragmatic cohort studies, which commonly have higher external validity compared to experimental randomized controlled trial studies (119). IMPR, however, could differ between different countries and continents, since rehabilitation centers around the globe work differently. Most of the scientific literature used in *Study I,* was based on data from rehabilitation centers from Western countries, such as USA, a number of European countries, Australia, and New Zealand, and the results of the systematic review/meta-analysis should therefore be extended to these countries, or to countries with similar conditions.

As relevant for *Study II* – results on inter-rater agreement – the raters were experienced and senior researchers in the field of pain rehabilitation, and the results should therefore be generalizable to raters with a similar specialist level. In addition, the results extend to QUIPS when judging the quality of the literature on prognostic factors, and with prompting items within the research field of pain rehabilitation. Our recommendations, however, could inspire researchers from other fields to work in a more structured manner, since much of the problems that were identified were rather generic and related to methodological issues such as poor reporting, multiple outcomes, and statistics. Moreover, since QUIPS is the recommended tool

to be used by the systematic reviews and meta-analyses in the Cochrane group, this study is of importance for evidence-based medicine, which is the base for general future rehabilitation recommendations for individual patient treatment.

6.6.2 Internal validity studies I-IV

6.6.2.1 *Study I*

The strength of this systematic review/meta-analysis is the broad approach that was used in order to synthesize the factors of importance for physical functioning, rather than exploring a single prognostic factor impact or a selected diagnosis within the chronic pain-population. The methodology used in this systematic review/meta-analysis follows strictly the standard recommendations and by the well-established Cochrane Prognosis Methods Group, and GRADE groups and the methodology used in the study has been pre-reviewed in a study protocol which was published before starting the analyses (79). Although the initial selection of papers based on title and abstract was mainly performed by only one reviewer, the screening process had a robust arrangement with randomization of studies and independent teams constituted by a senior and junior researcher (Figure 5). The broad selection of studies resulted in high heterogeneity, measured by I^2, for almost all comparisons (range: 48% to 94%), but there were no systematic reasons for this variance. Sensitivity analyses showed that the results were robust when controlling for different factors such as follow-up time, study quality, and statistical analyses. Although there was an unacceptably low inter-rater agreement for some RoB domains, it was not difficult to obtain consensus on the overall RoB scores.

One of the limitations of the systematic review/meta-analysis that could have occurred is the presence of *information bias*, since "gray" and non-English literature was omitted. In addition, the grouping of factors and outcomes, which were measured with various instruments could have increased *heterogeneity* and for that reason, incompatible and vague measures were excluded from the analyses. Another source for limitation of the study could be the *low study quality* of the included studies. However, it is not clear if the studies with low quality were actually biased, since in many of the studies there was a lack of relevant reporting on for example, study participation and attrition. A final point of limitation could be the occurrence of *publication bias*. The relatively small number of studies reporting on each comparison hampered the possibility for making detailed and meaningful analysis of funnel plots.

6.6.2.2 *Study II*

The use of systematic review and meta-analysis has increased over the years, since they are of the highest importance for evidence-based medicine, and issues of quality assessment have currently been structured in order to reach transparency in the process. One strength of this study is that not many previous studies have been published in the field of rehabilitation of chronic pain on experiences and learning processes of using methodological tools, and tried to improve such instruments, which is what this study has attempted to do.

The number of raters (and papers) included in our analyses was relatively small and this could be an important limitation of this study. This resulted in low prevalence in some of the cells, especially when performing sub-group analyses. Perhaps the inclusion of the third (junior) reviewer could have increased the inter-rater reliability to some extent. The lack of a "true RoB" for each paper could have led to an overestimation of the Kappa agreement if both raters made the same error.

6.6.2.3 *Study III*

A strength of this study was the large sample size that had national coverage. A relatively high proportion (47%) of patients were lost to follow-up, which is similar to earlier studies with long-term follow-up (84). A crude missingness analysis showed, however, that the non-responders were similar to the responders on all pain characteristic variables and self-rated health measures, but a larger proportion of the non-responders was born outside Europe and had a somewhat lower degree of vocational connection. This may indicate that information on study participation and questionnaires was not successfully adapted for foreign participants. Despite the width of tested potential prognostic factors, the degree of explained variance in final regression models did not exceed 21%, indicating that for the complete prediction of outcome, much still remains unrevealed.

6.6.2.4 *Study IV*

The independent variable Duration: Some programs at certain clinics are relatively stipulated, others permit more variability in the treatment according to the patient clinical status. Better 'recording', or register-documenting, of actual treatment is, however, needed, particularly since no standardized programs exist (45) with regard to content, duration, intensity, and adherence.

Since the independent variable on IMPR duration is an instrumental variable, theoretically the association between the dependent and independent variables cannot be influenced by confounders. Still, we tested if all pre-intervention data and the variable "treatment days per week" were related to the independent variable (treatment duration) and could have a potential to confound any associations between dependent and independent variables in the between-group analyses. Regression analyses showed though that a longer duration was somewhat related to "treatment days/week" ($r = -0.25$), indicating that the longer duration groups have less treatment days/week. This variable was therefore adjusted for in all our between-group GEE-analyses, as was all baseline data.

Our comparison groups were not randomized, and although our analysis allowed controlling for between-group differences, unknown potential confounding has not been controlled for.

6.7 STATISTICAL CONSIDERATIONS

Study I: Using the inverse variance method in the meta-analysis, the weight given to each study is the inverse of the variance of the effect estimate (i.e. one over the square of its standard

error). In this way larger studies are given more weight than smaller studies, which have larger standard errors. This choice of weight minimizes the uncertainty/ imprecision of the pooled effect estimate.

Studies III and IV: Discussion on using the Minimal clinically important difference as cut-off for improvement (*Study III*) and for improvement and deterioration (*Study IV*). Logistic regression models require cut-offs, for which a clear rationale is not always present. For the dichotomization of outcomes and the definitions used for improvement we used previously reported MCID definitions, however, to our knowledge these are not yet comprehensively validated in patients with chronic pain, which suggests some attention is required to what a good outcome represents using MCID (91). Also, in *Study IV*, we based the analyses on proportion of MCID improvement *and* deterioration on the assumption that MCID scores would not differ bidirectionally, which of course could be argued upon.

6.8 GENERAL DISCUSSION

In this thesis, three studies concerned various aspects of prognosis. In this respect, it is important to remember the distinction in interpretation between prognostic and predictive (120, 121), although a factor may be a little of both, the prognostic factor indicates the likely outcome (for instance disease recurrence) *irrespective* of treatment, whereas predictive refers to the likely benefit *after* treatment, compared to the patient's condition at baseline. Our bivariate outcome was created to reflect an improvement – and therefore also give some indication of the likely benefit interpreted as a result of the treatment. However, and here an important distinction is necessary, prognostic factors are not intended for treatment selection, and not suitable for interpretation as if those lacking the prognostic factor would benefit less from the treatment.

Taken together the prognostic factors revealed from *Studies I and III*, these studies point to a clinical picture, where the baseline prerequisites for an improved physical functioning – 'a better in better out' – here can be interpreted as to be able to work with behavior change in relation to exercise, activities, increasing function. One needs to be in 'pretty good shape' both physically and mentally as well as have access to adaptive coping-mechanisms. When it comes to mental functioning, results are based on *Study III* and show another clinical picture, indicating here that the person in worse shape is more likely to improve.

The outcome measures used in this thesis relate mostly to general measures of functioning, and to general life impact concepts such as HRQoL. This seems of relevance as this is an overall target of rehabilitation in general, and specifically in the particular case of chronic pain considering the low levels of quality of life reported for this population. Even though marginal effects were modest, hopefully they represent a perceived change of life impact of their pain condition. Considering the focus on physical and mental core domains in this thesis, it may be that the equally important third, social dimension, in this concept is not as well reflected in the way we assessed HRQoL.

Physical exercise and conditioning is definitely one of the cornerstones of IMPR practice. During this project we have come across empirical data that suggest there may be an imbalance in measures relating to physical functioning compared to emotional and psychosocial measures, for instance those reported in the SQRP. Psychosocial measures are evidently of greatest relevance, but where does all the information on physical fitness and strength training, i.e. the more proximal dimensions of physical functioning, end up? Is it possible that there is an unfortunate loss of information that is potentially relevant for our understanding of IMPR effectiveness?

With our chosen MCID-levels, at the least quite a large proportion improved over time. As indicated, however, some also experienced decreased ratings, equivalent to MCID levels, and some remained unchanged. Of course, one wants to go deeper into those who did not improve or who deteriorated. And also, we need to validate further whether these MCID levels were appropriate. Another issue that could be discussed is whether the reverted MCID were an appropriate measure?

All in all, the studies included in this thesis could be of importance to the existing evidence base, and in the long term they could provide guidance to all healthcare personnel working to make a difference to sufferers of chronic pain conditions.

6.8.1 Clinical implications

From the systematic review and from the large quality register, it was evident that measures of physical functioning were not as rich in detail as those of emotional functioning. Considering the comprehensive approach taken to physical conditioning and less interference with activities in IMPR, these domains seem under-evaluated, but could easily be implemented with the many new performance based ways to measure physical activity.

Modifiable factors, positive affect, treatment expectancy, hope, control, active coping mechanisms? All of these are surely already addressed – but perhaps with more focus on decreasing negative cognitive behavioral coping such as fear avoidance, in order to restore health. An interesting, complementary way, would be to increase sustainability by targeting protective cognitive behavioral factors, such as Acceptance and Commitment Therapy, which is gaining increasing ground. This might be an area where further gains can be achieved.

More data is needed on received actual intervention. A minimum level of information needs to be transformed to the SQRP, to enable analysis of progress in more detail.

For researchers using QUIPS, the following tips are highlighted. First, the importance of agreeing on the judgment of the prompting items in advance, an initial pilot on 2-3 articles can help identify areas in need of clarification. Secondly, plan for an early tuning-discussion with the team (after 5-10 articles) to help resolve dilemmas and disagreements that have arisen in

praxis. Finally, calculate Kappa early, as a way of identifying domains that may be in need of harmonization.

6.8.2 Future research

- Future prognostic research needs to investigate how personal and treatment factors *in combination* may explain more of the 'unexplained variance',
- Since some of the present baseline factors go in a 'better in better out' direction while others are 'the worse the better', presenting a complex prognostic picture, there is a need for further studies on what is behind such result for the complete understanding of good clinical follow-up,
- In addition to how prognostic variables coincide with the clinical landscape, with the help of the present results, prognostic models and clinical tools are now needed for how these variables can be used for adaptation to the *individual patient*. In connection with this, research is also needed on how such models can be *implemented* in the IMPR team work,
- There is a need for accessibility and knowledge to generate register-based randomized controlled trials that can be linked to SQRP. In this way, questions on different treatment approaches (e.g. IMPR intensity) can – I believe – be answered with clearer causality, and at relatively low costs,
- There is also a need for research collaborations that provide broad access to linked registry databases, that SQRP is linked to the social insurance database as well as to drug registers and e.g. registers that handle socio-demographic data. In this way, unanswered questions can be approached that relate more strongly to the patients' social situation and societal participation.
- The psychometrical properties of the outcome instrument used in SQRP needs more research,
- There is – I believe – a need to recognize physical activity patterns by using modern machine learning algorithms on activity tracker data, and by adding such objective data investigate what patterns predict clinical relevant improvements on clinical outcomes. That is since the research in the field of physical activity has been shown promising in various patient group suffering from musculoskeletal disorders.
- Finally, there is a great need to link the knowledge gained from PROMS with what can be measured at biomechanical laboratories on human movements in this major group. Such research is fundamental for how rehab exercises and physical activity should be adapted and balanced, and research related to these issues has begun as a result of this thesis work.

7 CONCLUSIONS

- Evidence for prognostic factors that precede IMPR was identified, providing suggestions for the targeting of modifiable factors in clinics and in future clinical trials. Specifically, there was a moderate to low level of evidence that better physical functioning after IMPR in patients with chronic pain is predicted by high levels of initial self-assessed physical functioning, low levels of emotional distress, low levels of cognitive and behavioral risk factors, and high levels of protective cognitive and behavioral factors, but not by initial pain level or pain duration (chronicity).

- Inter-rater agreement using the Quality in Prognosis Studies-tool improved over time and despite relatively weak inter-rater agreement, this proved to be a useful tool for assessing the risk of bias when performing a meta-analysis of prognostic studies in pain rehabilitation. The main strength of this tool is that it requires raters to discuss and investigate important aspects of study quality. Hence, the quality assessment of published results needs systematic consensus work between assessors.

- Employment status, younger age, high beliefs of restored health, high emotional health, low levels of pain, and pain interference, but high levels of anxiety and worse initial physical health all predicted *Physical functioning. Emotional functioning*, on the other hand, was predicted by employment status, high general activity, high sense of life control, and of European origin, but worse initial emotional health, indicating a complex prognostic picture for the complete understanding of good clinical follow-up.

- The effectiveness of IMPR programs for patients with chronic pain was supported across all outcomes in the biopsychosocial spectrum in the short term (post-intervention) and long term. Still, there was a noticeable proportion of patients who deteriorated during the follow-up period.

- Despite some statistical differences favoring moderate-duration IMPR, no clinically meaningful differences emerged for comparative effectiveness across all outcomes, which suggests that the IMPR content and team collaborative aspects can be highlighted in future rehabilitation developments.

8 ACKNOWLEDGEMENTS

My time as a PhD student at Karolinska Institutet, Division of Physiotherapy, my workplace for so many years, has been a very stimulating experience and a tremendous challenge. Along the way I have been rewarded with so much help and support from many skillful, helpful and friendly people who have helped me on this journey, and now I have arrived at my destination. For those of you who know me, you know this is the hardest part, but I'll try to make it short and I would like to express my deepest gratitude and appreciation to all of you, and in particular to the following:

Björn Äng, my main supervisor, thank you, for everything! First of all, for providing me with the opportunity to become your PhD student – I will always be grateful for that. And then for everything else…I think you've been a fantastic supervisor, combining scientific rigor with your unique Björn Äng spirit. I've learned so much from you, and I'm still trying to. Thank you!

Wim Grooten, my co-supervisor. Thank you too, for just about *everything!* Your generosity and willingness to help, and that you are always close to laughter. Thank you for all the diamonds, and we both know 'more is more' but in this case one is enough, and that is you.

Björn Gerdle, my co-supervisor. Thank you for kindly sharing your knowledge, giving encouragement and providing support throughout the years. Your prestigeless attitude towards your expertise, especially during the systematic review process – I've enjoyed all the discussions we've had so much.

The Department and Department Heads
The Department of Neurobiology, Care Sciences and Society, Division of Physiotherapy, Karolinska Institutet, for the opportunity to become a doctoral student.

The former and present heads of the Division of Physiotherapy: Cecilia Fridén, Annette Heijne, Maria Hagströmer, Malin Nygren Bonnier and Joanna Kvist. Thank you all for your support, encouragement and belief in me, and solid back up when most needed. A special thanks to Annette, my mentor in hard times, for sharing so generously with your characteristic warmth and wisdom.

My co-authors from the systematic review project: Britt-Mari Stålnacke, Katja Boersma and Paul Enthoven, for sharing your excellent knowledge with me, and for taking the leap with me, boarding the ship of infinite systematic review. I have learned so much from our discussions – it was an extra bonus luxury!

Riccardo Lo Martire, co-author and fellow doctoral student, virtuoso in mathematics and statistics, for introducing me to everything around 'everything that is complicated'.

Linda Vixner, my co-author and inspirational role model in oh, too many ways to mention here.

The Research group at the Division of Physiotherapy, Musculoskeletal disorders from a biopsychosocial perspective, thank you for all good times and fruitful and fun discussions.

Högskolan Dalarna, and REPS in particular for providing such a good work environment. A special thanks to my research group and fellow doctoral students in Falun Veronica and Jens, and Lena Pettersson for keen administrative help. Thank you all for making my trips to Falun so pleasant.

Moudud Alam, Associate professor and statistician at Högskolan Dalarna, for your kind support with statistical queries and your gentle ways of explaining so may complicated things and make them apprehensible.

Gabrielle Mackay Thomsson, master of the arts in language editing and in refining the text to read so beautifully and clearly. How can I *ever* thank you for all your fine work?

Klas Moberg, research librarian at University Library of Karolinska Institutet,you're your knowledgeable and kind support with the preparation of the search strategies for the systematic review in Study I. I learned so much from you!

All the people involved in the SQRP; A special thanks to all units who work with providing all the data to the register; to Elisabeth Persson, Anneli Inghilesi, Marcelo Rivano Fischer for providing all the additional support when needed. Also my thanks extend to the SQRP Research network and the SQRP Doctoral students group – both of which have been very nourishing for my understanding of the field of rehabilitation in chronic pain.

My very own 'pain expertise group', constituted by my first inspirational models within the pain field and with whom I've shared so many discussions, on the course to developing the Smärtkursen at the Physiotherapy program; Iréne Lund, mentor and friend for so long now, Monika Löfgren and Cecilia Norrbrink, also mentors and friends. You can't have too many.

Smärtsektionen – thank you all for sharing you expertise from the field.

Colleagues
ALL my colleagues at the Division of Physiotherapy. My long-standing teaching colleagues and friends – it's all mixed into one; Annica, Charlotte, Ann-Mari, Birgitta, Christina, Veronica and ALL OF YOU. You're the best!

And a special thanks to Vanja Landin, Balbir Dhuper, Kathrin Dellblad, Ellinor Nordström and Martina Bergenroth for always helping out with everything around practicalities with doctoral studies and those infinite questions.

Present and former doctoral students, with whom I've shared this journey
My friends in the National Postgraduate School of Health Care Sciences, for all the good times. Special thanks go to Annika, Susanne, Mia and Henrik – I'm grateful for everything we've shared, mixing serious discussion with laughter, through good times and hard times, and for all statistic lovers.

All my colleagues in the A3 corridor; who make that place such a wonderful place to be, yes Kirsti I'm talking about you, and of course all of you others too, but you know what I mean. Breiffni, thank you for helping me with everything.

Special thanks also goes to all of you with whom I've shared the doctoral student's room for the past four years: Thomas, Håkan, Andreas, Conran, Ingrid, Emelie BF, Philip, Emelie K and Tobias, Sebastian, Malin, Anna, Andrea and John. I'm truly happy for the times we've spent together – and each and everyone of you, I owe sincere gratefulness and many happy memories from good discussions. And Andreas, thanks for saving my life. My heart aches that I have to leave the doctoral student's room now. Can we not revise the rules?? A special thanks to Hanna and Emelie – spending time with you has added so much value to my time as a doctoral student, and because of you I feel younger (on the inside).

Actually, apart from the times with great agony, there were a few, at the same time I don't know if I've ever laughed this much as I have in the past four years.

There are also many friends from outside KI, thank you for staying put and having faith in me even if I've been away for long now. My best friend from childhood, Yukiko- what can I say- your're always in my heart, a true friend from the very first day you bumped into me in the pulkabacken, and all of the family is my family as well, true "cousins".

My family, for always believing in me, no matter what. My mother, Maria, my father Bengt, Marie-Louise, my sisters Katarina and Ulrika with families and Thanasis and Kostas – I'm blessed. I've been blessed with so much love and support around me, and I am truly the luckiest person, surrounded only by loving people – don't ever stop being who you are. And I look so much forward to seeing you all more again.

Giorgos, love of my life and steady rock to lean on. I don't know what to say, I couldn't have done this without you, thank you so much for everything that you are.

Ioannis, joy of my life. Du är det finaste som finns.

Oh, I get by with a little help from my friends
with a little help from my friends…

9 REFERENCES

1. Mahler H. Health for all by the year 2000 WHO Chron. 1975(29):457-61.
2. Organization WH. Preamble to the Constitution of WHO as adopted by the International Health Conference, New York, 19 June - 22 July 1946; signed on 22 July 1946 by the representatives of 61 States (Official Records of WHO, no. 2, p. 100) and entered into force on 7 April 1948. . In: Organization WH, editor.
3. WHO. International Classification of Functioning, Disability and Health (ICF). Geneva: World Health Organisation2001.
4. McDougall J, Wright V, Rosenbaum P. The ICF model of functioning and disability: incorporating quality of life and human development. Dev Neurorehabil. 2010;13(3):204-11.
5. Dworkin RH, Turk DC, Farrar JT, Haythornthwaite JA, Jensen MP, Katz NP, et al. Core outcome measures for chronic pain clinical trials: IMMPACT recommendations. Pain. 2005;113(1-2):9-19.
6. Merskey H, Bogduk N, International Association for the Study of Pain. Task Force on T. Classification of chronic pain : descriptions of chronic pain syndromes and definitions of pain terms. 2. ed. ed. Seattle: Seattle : IASP Press; 1994.
7. IASP. IASP Terminology: International Association for the Study of Pain; 2017 [Available from: https://www.iasp-pain.org/Education/Content.aspx?ItemNumber=1698&navItemNumber=576.
8. Williams AC, Craig KD. Updating the definition of pain. 2016(1872-6623 (Electronic)).
9. Kosek E, Cohen M, Baron R, Gebhart GF, Mico J-A, Rice ASC, et al. Do we need a third mechanistic descriptor for chronic pain states? Pain. 2016;157(7).
10. Arendt- Nielsen L, Morlion B, Perrot S, Dahan A, Dickenson A, Kress H, et al. Assessment and manifestation of central sensitisation across different chronic pain conditions. European Journal of Pain. 2018;22(2):216-41.
11. Raffaeli W, Arnaudo E. Pain as a disease: an overview. J Pain Res. 2017;10:2003-8.
12. Nicholas M, Vlaeyen JWS, Rief W, Barke A, Aziz Q, Benoliel R, et al. The IASP classification of chronic pain for ICD-11: chronic primary pain. Pain. 2019;160(1):28-37.
13. Treede R-Da, Rief Wb, Barke Ab, Aziz Qc, Bennett MId, Benoliel Re, et al. A classification of chronic pain for ICD-11. Pain. 2015;156(6):1003-7.
14. Perrot S, Cohen M, Barke A, Korwisi B, Rief W, Treede RD, et al. The IASP classification of chronic pain for ICD-11: chronic secondary musculoskeletal pain. Pain. 2019;160(1):77-82.
15. Leadley RM, Armstrong N, Lee YC, Allen A, Kleijnen J. Chronic diseases in the European Union: the prevalence and health cost implications of chronic pain. J Pain Palliat Care Pharmacother. 2012;26(4):310-25.
16. Breivik H, Collett B, Ventafridda V, Cohen R, Gallacher D. Survey of chronic pain in Europe: prevalence, impact on daily life, and treatment. Eur J Pain. 2006;10(4):287-333.

17. van Tulder MW, Koes BW, Bouter LM. A cost-of-illness study of back pain in The Netherlands. Pain. 1995;62(2):233-40.

18. Vos T, Abajobir AA, Abate KH, Abbafati C, Abbas KM, Abd-Allah F, et al. Global, regional, and national incidence, prevalence, and years lived with disability for 328 diseases and injuries for 195 countries, 1990–2016: a systematic analysis for the Global Burden of Disease Study 2016. The Lancet. 2017;390(10100):1211-59.

19. Da Silva JAP, Geenen R, Jacobs JWG. Chronic widespread pain and increased mortality: biopsychosocial interconnections. Ann Rheum Dis. 2017.

20. Torrance N, Elliott AM, Lee AJ, Smith BH. Severe chronic pain is associated with increased 10 year mortality. A cohort record linkage study. European journal of pain. 2010;14(4):380-6.

21. SBU. Rehabilitering vid långvarig smärta: En systematisk litteraturöversikt: partiell uppdatering och fördjupning av SBU-rapport nr 177/1+2. Stockholm: Statens beredning för medicinsk utvärdering (SBU); 2010.

22. Försäkringskassan S. Vad kostar olika sjukdomar i sjukförsäkringen? 2011.

23. Arendt-Nielsen L, Fernández-de-Las-Peñas C, Graven-Nielsen T. Basic aspects of musculoskeletal pain: from acute to chronic pain. The Journal of manual & manipulative therapy. 2011;19(4):186-93.

24. Nicholas MK. Why do some people develop chronic, treatment-resistant pain and not others? Pain. 2018;159(12):2419-20.

25. Norrbrink C, Lundeberg T. (red). Om smärta : ett fysiologiskt perspektiv. 2. uppl. ed. Norrbrink C, Lundeberg T, editors. Lund: Lund : Studentlitteratur; 2014.

26. Melzack R, Wall PD. Pain Mechanisms: A New Theory. Science. 1965;150(3699):971.

27. Melzack R. From the gate to the neuromatrix. Pain. 1999;82.

28. Borg J. Rehabiliteringsmedicin : [teori och praktik]. 1. uppl. ed. Lund: Lund : Studentlitteratur; 2015.

29. Latremoliere A, Woolf CJ. Central Sensitization: A Generator of Pain Hypersensitivity by Central Neural Plasticity. The Journal of Pain. 2009;10(9):895-926.

30. Gatchel RJ, Peng YB, Peters ML, Fuchs PN, Turk DC. The biopsychosocial approach to chronic pain: scientific advances and future directions. Psychol Bull. 2007;133(4):581-624.

31. Boersma K, Linton SJ. Expectancy, fear and pain in the prediction of chronic pain and disability: a prospective analysis. European Journal of Pain. 2006;10(6):551-.

32. Laisne F, Lecomte C, Corbiere M. Biopsychosocial predictors of prognosis in musculoskeletal disorders: a systematic review of the literature (corrected and republished) *. Disabil Rehabil. 2012;34(22):1912-41.

33. Valentin GH, Pilegaard MS, Vaegter HB, Rosendal M, Ørtenblad L, Væggemose U, et al. Prognostic factors for disability and sick leave in patients with subacute non-malignant pain: a systematic review of cohort studies. BMJ Open. 2016;6(1).

34. van Hecke O, Torrance N, Smith BH. Chronic pain epidemiology and its clinical relevance. Br J Anaesth.111(1):13-8.

35. Linton SJ, Boersma K. Early identification of patients at risk of developing a persistent back problem: the predictive validity of the Örebro Musculoskeletal Pain Questionnaire. The Clinical journal of pain. 2003;19(2):80-6.

36. Engel GL. The Need for a New Medical Model: A Challenge for Biomedicine. Science. 1977;196(4286):129-36.

37. Gatchel RJ. The Continuing and Growing Epidemic of Chronic Low Back Pain. Healthcare (Basel). 2015;3(3):838-45.

38. Turk DC, Melzack R. Handbook of pain assessment. 3. ed. ed. New York: New York : Guilford Press; 2011.

39. Andersen LN, Kohberg M, Juul-Kristensen B, Herborg LG, Sogaard K, Roessler KK. Psychosocial aspects of everyday life with chronic musculoskeletal pain: A systematic review. Scand J Pain. 2014;5(2):131-48.

40. Dueñas M, Ojeda B, Salazar A, Mico JA, Failde I. A review of chronic pain impact on patients, their social environment and the health care system. J Pain Res. 2016;9:457.

41. Vartiainen P, Heiskanen T, Sintonen H, Roine RP, Kalso E. Health-related quality of life and burden of disease in chronic pain measured with the 15D instrument. Pain. 2016;157(10):2269-76.

42. Pless M, Granlund M. Handbok i att använda ICF och ICF-CY. Stockholm: Studentlitteratur; 2011.

43. Denison E. Beteendemedicinska tillämpningar i sjukgymnastik. 1. uppl. ed. Åsenlöf P, editor. Lund: Lund : Studentlitteratur; 2012.

44. Gerdle B, Stålnacke B-M, Söderlund A, Åsenlöf P. Indikation för multimodal rehabilitering vid långvarig smärta. 2011.

45. Kaiser U, Treede R-D, Sabatowski R. Multimodal pain therapy in chronic noncancer pain—gold standard or need for further clarification? Pain. 2017;158(10):1853–9.

46. Gatchel RJ, McGeary DD, McGeary CA, Lippe B. Interdisciplinary chronic pain management: past, present, and future. Am Psychol. 2014;69(2):119-30.

47. Scascighini L, Toma V, Dober-Spielmann S, Sprott H. Multidisciplinary treatment for chronic pain: a systematic review of interventions and outcomes. Rheumatology (Oxford). 2008;47(5):670-8.

48. Guzman J, Esmail R, Karjalainen K, Malmivaara A, Irvin E, Bombardier C. Multidisciplinary rehabilitation for chronic low back pain: systematic review. BMJ. 2001;322(7301):1511-6.

49. Craig P, Dieppe P, Macintyre S, Michie S, Nazareth I, Petticrew M. Developing and evaluating complex interventions: the new Medical Research Council guidance2008 2008-09-29 10:52:26.

50. Waterschoot FPC, Dijkstra PU, Hollak N, de Vries HJ, Geertzen JHB, Reneman MF. Dose or content? Effectiveness of pain rehabilitation programs for patients with chronic low back pain: A systematic review. PAIN®. 2014;155(1):179-89.

51. Bandura A, Walters RH. Social learning theory: Prentice-hall Englewood Cliffs, NJ; 1977.

52. Bovend'Eerdt TJ, Botell RE, Wade DT. Writing SMART rehabilitation goals and achieving goal attainment scaling: a practical guide. Clin Rehabil. 2009;23(4):352-61.

53. Stenberg G, Pietilä Holmner E, Stålnacke B-M, Enthoven P. Healthcare professional experiences with patients who participate in multimodal pain rehabilitation in primary care–a qualitative study. Disabil Rehabil. 2016;38(21):2085-94.

54. Turk DC, Dworkin RH, Allen RR, Bellamy N, Brandenburg N, Carr DB, et al. Core outcome domains for chronic pain clinical trials: IMMPACT recommendations. Pain. 2003;106(3):337-45.

55. Turk DC, Dworkin RH, Revicki D, Harding G, Burke LB, Cella D, et al. Identifying important outcome domains for chronic pain clinical trials: an IMMPACT survey of people with pain. Pain. 2008;137(2):276-85.

56. Taylor AM, Phillips K, Patel KV, Turk DC, Dworkin RH, Beaton D, et al. Assessment of physical function and participation in chronic pain clinical trials: IMMPACT/OMERACT recommendations. Pain. 2016;157(1872-6623 (Electronic)):1836–50.

57. Riley RD, Hayden JA, Steyerberg EW, Moons KGM, Abrams K, Kyzas PA, et al. Prognosis Research Strategy (PROGRESS) 2: Prognostic Factor Research. PLoS Med. 2013;10(2):e1001380.

58. de Rooij A, Roorda LD, Otten RH, van der Leeden M, Dekker J, Steultjens MP. Predictors of multidisciplinary treatment outcome in fibromyalgia: a systematic review. Disabil Rehabil. 2013;35(6):437-49.

59. van der Hulst M, Vollenbroek-Hutten MM, IJzerman MJ. A systematic review of sociodemographic, physical, and psychological predictors of multidisciplinary rehabilitation—or, back school treatment outcome in patients with chronic low back pain. Spine (Phila Pa 1976). 2005;30(7):813-25.

60. Wessels T, van Tulder M, Sigl T, Ewert T, Limm H, Stucki G. What predicts outcome in non-operative treatments of chronic low back pain? A systematic review. Eur Spine J. 2006;15(11):1633-44.

61. Kamper SJ, Apeldoorn AT, Chiarotto A, Smeets RJEM, Ostelo RWJG, Guzman J, et al. Multidisciplinary biopsychosocial rehabilitation for chronic low back pain: Cochrane systematic review and meta-analysis. BMJ. 2015;350.

62. SBU. Metoder för behandling av långvarig smärta. En systematisk litteraturöversikt/ Vol 1+2. Stockholm: Stockholm : Statens beredning för medicinsk utvärdering (SBU); 2006.

63. van Middelkoop M, Rubinstein SM, Kuijpers T, Verhagen AP, Ostelo R, Koes BW, et al. A systematic review on the effectiveness of physical and rehabilitation interventions for chronic non-specific low back pain. Eur Spine J. 2011;20(1):19-39.

64. Haldorsen EMH, Grasdal AL, Skouen JS, Risa AE, Kronholm K, Ursin H. Is there a right treatment for a particular patient group? Comparison of ordinary treatment, light multidisciplinary treatment, and extensive multidisciplinary treatment for long-term sick-listed employees with musculoskeletal pain. Pain. 2002;95(1-2):49-63.

65. Richardson J, Loyola-Sanchez A, Sinclair S, Harris J, Letts L, MacIntyre NJ, et al. Self-management interventions for chronic disease: a systematic scoping review. Clin Rehabil. 2014;28(11):1067-77.

66. Sundell K. Om evidensbaserad praktik: Socialstyrelsen; 2012.

67. Murad MH, Asi N, Alsawas M, Alahdab F. New evidence pyramid. Evidence Based Medicine. 2016;21(4):125.

68. Aschengrau A, Seage GR. Essentials of epidemiology in public health. Sudbury, Mass. [u.a.]: Jones and Bartlett Publ.; 2008.

69. Hayden JA, van Der Windt DA, Cartwright JL, Côté P, Bombardier C. Assessing bias in studies of prognostic factors. Ann Intern Med. 2013;158(4):280.

70. Guyatt GH, Oxman AD, Kunz R, Vist GE, Falck-Ytter Y, Schünemann HJ. What is "quality of evidence" and why is it important to clinicians? BMJ. 2008;336(7651):995.

71. Whiteneck GG, Gassaway J. SCIRehab Uses Practice-Based Evidence Methodology to Associate Patient and Treatment Characteristics With Outcomes. Arch Phys Med Rehabil. 2013;94(4):S67-S74.

72. Hernán MA, Robins JM. Using Big Data to Emulate a Target Trial When a Randomized Trial Is Not Available. Am J Epidemiol. 2016;183(8):758-64.

73. Emilsson L, Lindahl B, Köster M, Lambe M, Ludvigsson JF. Review of 103 Swedish Healthcare Quality Registries. J Intern Med. 2015;277(1):94-136.

74. Adami HO, Hernán MA. Learning how to improve healthcare delivery: the Swedish Quality Registers. J Intern Med. 2015;277(1):87-9.

75. Sveriges Kommuner och Landsting S. Nationella Kvalitetsregister 2019 [Available from: http://kvalitetsregister.se/.

76. Nationella Registret över Smärtrehabilitering N. NRS - Nationella Registret över Smärtrehabilitering 2019 [updated 2019-02-05. Available from: https://www.ucr.uu.se/nrs/om-registret.

77. Sveriges Kommuner och Landsting S. Find a registry: National Quality Registry for Pain Rehabilitation (NRS) 2019 [Available from: http://www.kvalitetsregister.se/englishpages/findaregistry/registerarkivenglish/nationalqualityregistryforpainrehabilitationnrs.2246.html.

78. Kaiser U, Kopkow C, Deckert S, Neustadt K, Jacobi L, Cameron P, et al. Developing a core outcome domain set to assessing effectiveness of interdisciplinary multimodal pain therapy: The VAPAIN consensus statement on core outcome domains. Pain. 2018;159(4):673-83.

79. Tseli E, Grooten WJA, Stålnacke B-M, Boersma K, Enthoven P, Gerdle B, et al. Predictors of multidisciplinary rehabilitation outcomes in patients with chronic musculoskeletal pain: protocol for a systematic review and meta-analysis. Systematic reviews. 2017;6(1):199.

80. Cooper HM. Research synthesis and meta-analysis : a step-by-step approach. 4th ed. ed. Cooper HM, editor. Los Angeles, [Calif.]: London: SAGE; 2010.

81. SBU. Utvärdering av metoder i hälso-och sjukvården: en handbok. 2 uppl. 2nd ed. (SBU) Sbfmu, editor. Stockholm: Stockholm: Statens beredning för medicinsk utvärdering SBU; 2014.

82. Covidence. Covidence systematic review software. Melbourne, Australia: Vertitas Health Innovation Ltd.

83. Moher D, Liberati A, Tetzlaff J, Altman DG, Group P. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. BMJ. 2009;339:b2535.

84. Tseli E, Boersma K, Stålnacke B-M, Enthoven P, Gerdle B, Äng BO, et al. Prognostic Factors for Physical Functioning After Multidisciplinary Rehabilitation in Patients With Chronic Musculoskeletal Pain: A Systematic Review and Meta-Analysis. The Clinical journal of pain. 2019;35(2):148-73.

85. Review Manager (RevMan) Version 5.3. ed. Copenhagen: The Nordic Cochrane Centre: The Cochrane Collaboration; 2014.

86. Cooper HM, Hedges LV, Valentine JC. The handbook of research synthesis and meta-analysis. 2nd ed. / edited by Harris Cooper, Larry V. Hedges and Jeffrey C. Valentine. ed. New York: New York : Russell Sage Foundation; 2009.

87. Atkins D BD, Briss PA, Eccles M, Falck-Ytter Y, Flottorp S, Guyatt GH, Harbour RT, Haugh MC, Henry D, Hill S, Jaeschke R, Leng G, Liberati A, Magrini N, Mason J, Middleton P, Mrukowicz J, O'Connell D, Oxman AD, Phillips B, Schünemann HJ, Edejer T, Varonen H, Vist GE, Williams JW Jr, Zaza S; GRADE Working Group. Grading quality of evidence and strength of recommendations. BMJ. 2004;328(1756-1833 (Electronic)):1490.

88. Huguet A, Hayden JA, Stinson J, McGrath PJ, Chambers CT, Tougas ME, et al. Judging the quality of evidence in reviews of prognostic factor research: adapting the GRADE framework. Systematic reviews. 2013;2:71.

89. Edwards RR, Dworkin RH, Turk DC, Angst MS, Dionne R, Freeman R, et al. Patient phenotyping in clinical trials of chronic pain treatments: IMMPACT recommendations. Pain. 2016;157(9):1851-71.

90. Auffinger BM, Lall RR, Dahdaleh NS, Wong AP, Lam SK, Koski T, et al. Measuring surgical outcomes in cervical spondylotic myelopathy patients undergoing anterior cervical discectomy and fusion: assessment of minimum clinically important difference. PLoS One. 2013;8(6):e67408.

91. Lauche R, Langhorst J, Dobos GJ, Cramer H. Clinically meaningful differences in pain, disability and quality of life for chronic nonspecific neck pain–a reanalysis of 4 randomized controlled trials of cupping therapy. Complement Ther Med. 2013;21(4):342-7.

92. Farrar JT, Young Jr JP, LaMoreaux L, Werth JL, Poole RM. Clinical importance of changes in chronic pain intensity measured on an 11-point numerical pain rating scale. Pain. 2001;94(2):149-58.

93. Salaffi F, Stancati A, Silvestri CA, Ciapetti A, Grassi W. Minimal clinically important changes in chronic musculoskeletal pain intensity measured on a numerical rating scale. European journal of pain. 2004;8(4):283-91.

94. McKillop JM, Nielson WR. Improving the usefulness of the Multidimensional Pain Inventory. Pain Research and Management. 2011;16(4):239-44.

95. Dworkin RH, Turk DC, Wyrwich KW, Beaton D, Cleeland CS, Farrar JT, et al. Interpreting the Clinical Importance of Treatment Outcomes in Chronic Pain Clinical Trials: IMMPACT Recommendations. The Journal of Pain. 2008;9(2):105-21.

96. Puhan MA, Frey M, Büchi S, Schünemann HJ. The minimal important difference of the hospital anxiety and depression scale in patients with chronic obstructive pulmonary disease. Health and quality of life outcomes. 2008;6(1):46.

97. Coretti S, Ruggeri M, McNamee P. The minimum clinically important difference for EQ-5D index: a critical review. Expert Rev Pharmacoecon Outcomes Res. 2014;14(2):221-33.

98. Lipsey MW, Wilson D. Practical meta-analysis (applied social research methods)2000.

99. Lenhard WL, A. . Calculation of Effect Sizes. Dettelbach (Germany): Psychometrica.; 2016.

100. Wilson DB, Ph.D. (n.d.). . Practical Meta-Analysis Effect Size Calculator. [Online calculator]. . The Campbell Collaboration; 2015.

101. Egger M. Systematic Reviews in Health Care : Meta-Analysis in Context. 2 ed. Chichester, GB: BMJ Books; 2008.

102. Lowry R. VassarStats: Website for Statistical Computation 2019 [Available from: http://vassarstats.net/.

103. MedCalc Software 2019 [Available from: https://www.medcalc.org/calc/comparison_of_proportions.php.

104. Zetterqvist J, Vansteelandt S, Pawitan Y, Sjölander A. Doubly robust methods for handling confounding by cluster. Biostatistics. 2016;17(2):264-76.

105. Sjölander A. Regression standardization with the R package stdReg. Eur J Epidemiol. 2016;31(6):563-74.

106. Jaeschke R, Singer J, Guyatt GH. Measurement of health status: ascertaining the minimal clinically important difference. Control Clin Trials. 1989;10(4):407-15.

107. Grooten WJA, Tseli E, Äng BO, Boersma K, Stålnacke B-M, Gerdle B, et al. Elaborating on the assessment of the risk of bias in prognostic studies in pain rehabilitation using QUIPS—aspects of interrater agreement. Diagnostic and Prognostic Research. 2019;3(1):5.

108. Network BB. On medicine [Internet]. Network BB, editor2019.

109. Heiskanen T, Roine RP, Kalso E. Multidisciplinary pain treatment–Which patients do benefit? Scandinavian journal of pain. 2012;3(4):201-7.

110. Vartiainen P, Heiskanen T, Sintonen H, Roine RP, Kalso E. Health- related quality of life change in patients treated at a multidisciplinary pain clinic. European Journal of Pain. 2019.

111. Rivano Fischer M, Persson EB, Stalnacke BM, Schult ML, Lofgren M. Return to work after interdisciplinary pain rehabilitation: One- and two-year follow-up based on the Swedish Quality Registry for Pain rehabilitation. J Rehabil Med. 2019;51(4):281-9.

112. Gerdle B, Akerblom S, Brodda Jansen G, Enthoven P, Ernberg M, Dong HJ, et al. Who benefits from multimodal rehabilitation - an exploration of pain, psychological distress, and life impacts in over 35,000 chronic pain patients identified in the Swedish Quality Registry for Pain Rehabilitation. J Pain Res. 2019;12:891-908.

113. Gerdle B, Molander P, Stenberg G, Stalnacke BM, Enthoven P. Weak outcome predictors of multimodal rehabilitation at one-year follow-up in patients with chronic pain-a practice based evidence study from two SQRP centres. BMC Musculoskelet Disord. 2016;17(1):490.

114. Goubert L, Trompetter H. Towards a science and practice of resilience in the face of pain. Eur J Pain. 2017;21(8):1301-15.

115. Hayden JA, Côté P, Bombardier C. Evaluation of the quality of prognosis studies in systematic reviews. Ann Intern Med. 2006;144(6):427-37.

116. Ringqvist A, Dragioti E, Bjork M, Larsson B, Gerdle B. Moderate and Stable Pain Reductions as a Result of Interdisciplinary Pain Rehabilitation-A Cohort Study from the Swedish Quality Registry for Pain Rehabilitation (SQRP). J Clin Med. 2019;8(6).

117. Volker G, van Vree F, Wolterbeek R, van Gestel M, Smeets R, Köke A, et al. Long-Term Outcomes of Multidisciplinary Rehabilitation for Chronic Musculoskeletal Pain. Musculoskeletal Care. 2017;15(1):59-68.

118. Dragioti E, Björk M, Larsson B, Gerdle B. A Meta-Epidemiological Appraisal of the Effects of Interdisciplinary Multimodal Pain Therapy Dosing for Chronic Low Back Pain. Journal of Clinical Medicine. 2019;8(6):871.

119. Iorio A, Spencer FA, Falavigna M, Alba C, Lang E, Burnand B, et al. Use of GRADE for assessment of evidence about prognosis: rating confidence in estimates of event rates in broad categories of patients. BMJ : British Medical Journal. 2015;350.

120. Ruberg SJ, Shen L. Personalized Medicine: Four Perspectives of Tailored Medicine. Statistics in Biopharmaceutical Research. 2015;7(3):214-29.

121. Sechidis K, Papangelou K, Metcalfe PD, Svensson D, Weatherall J, Brown G. Distinguishing prognostic and predictive biomarkers: an information theoretic approach. Bioinformatics. 2018:bty357-bty.